THE FLORIDA STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

EVIDENCE FOR SINGLE TOP QUARK PRODUCTION USING BAYESIAN NEURAL NETWORKS

By

DAEKWANG KAU

A Dissertation submitted to the
Department of Physics
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Degree Awarded:
Fall Semester, 2007

The members of the Committee approve the Dissertation of Daekwang Kau defended on August 20, 2007.

Harrison B. Prosper
Professor Directing Dissertation

Ettore Aldrovandi
Outside Committee Member

Todd Adams
Committee Member

Laura Reina
Committee Member

Jorge Piekarewicz
Committee Member

Approved:

Mark Riley, Chair
Department of Physics

Joseph Travis, Dean, College of Arts and Sciences

The Office of Graduate Studies has verified and approved the above named committee members.

This thesis is dedicated to Yun Jeong, Jennifer, and my parents

ACKNOWLEDGEMENTS

First of all, I would like to acknowledge my advisor and mentor, Dr. Harrison B. Prosper, for his guidance and support throughout my graduate career. Without his support, I could not have finished the research and dissertation. I am also grateful to the members of my committee, Dr. Ettore Aldrovandi, Dr. Todd Adams, Dr. Laura Reina, and Dr. Jorge Piekarewicz.

Special thanks go to current and previous convenors of the singletop working group, Dr. Ann Heinson, Dr. Aran Garcia-Bellido, and Dr. Reinhard Schwienhorst for providing opportunities to work on this wonderful physics topic, to Dr. Supriya Jain for helping me to understand statistical issues, and to all singletop group members for useful discussions.

I would like to thank the professors, postdocs, and students in the high energy physics (HEP) group of Florida state university (FSU). I had wonderful time with Sinjini (now a research associate at the University of Minnesota) and Haryo in and also out of the DØ trailer.

I have enjoyed working with Dr. Kazunori Hanagaki on the development of the layer 0 detector. I learned many useful technical skills during the period I worked with him. Also, I would like to thank him for recommendation letters.

When I first came to Fermilab, I was enormously helped by Dr. Dookee Cho and Sunghwan Ahn. Without them, I would have gone through more trial and error.

Last but not least, I am thankful to my family and friends for believing in me and supporting me. I feel I am particularly indebted to my parents raising my baby. Even if they have sacrificed their time, I hope they are proud of me.

TABLE OF CONTENTS

List of Tables vii

List of Figures viii

Abstract xiv

Preface 1

1. Introduction 2
- 1.1 Standard Model 2
- 1.2 The Top Quark 10
- 1.3 Single Top Quark 12

2. Accelerator and Detector 16
- 2.1 Tevatron Accelerator 16
- 2.2 DØ Detector 19

3. Event Reconstruction 32
- 3.1 Tracks 32
- 3.2 Primary Vertices 34
- 3.3 Electrons 34
- 3.4 Muons 35
- 3.5 Jets 37
- 3.6 b Quark Jets 39
- 3.7 Missing Transverse Energy 40

4. Data Samples 41

5. Signal and Background Estimation 45
- 5.1 Single Top Event Generation 45
- 5.2 Modeling Backgrounds 45
- 5.3 Background Generation 46
- 5.4 Trigger Simulation 48
- 5.5 Monte Carlo Corrections 50

6. Event Selection 54

7. Bayesian Neural Networks . . . 57
7.1 Basic Concepts of Neural Networks . . . 58
7.2 Bayes' Theorem . . . 60
7.3 Bayesian Neural Networks . . . 61
7.4 Markov Chain Monte Carlo (MCMC) . . . 64

8. Analysis . . . 72
8.1 Introduction . . . 72
8.2 Variable Selection . . . 75
8.3 BNN Training and Verification . . . 83
8.4 BNN Outputs . . . 98
8.5 Cross Check Samples . . . 100
8.6 Measuring Single Top Cross Section . . . 102
8.7 Probabilities and Signal Significance . . . 112
8.8 Summary and Conclusions . . . 114
Appendices . . . 115

A. Discrepancy Measures . . . 115

B. Prior Width Estimates Using JETNET . . . 119

C. Plots After Selection . . . 122

References . . . 128

Biographical Sketch . . . 134

LIST OF TABLES

4.1 Integrated luminosities by trigger version for the electron channel. 41

4.2 Integrated luminosities by trigger version for the muon channel. 42

5.1 Event numbers, cross section, and weights for $t\bar{t}$ MC events. 46

5.2 Event numbers, cross section, and weights for W+jets MC events. 48

6.1 Numbers of events for the electron and muon channels after selection. 56

8.1 Initial list of variables, in three categories: object kinematics; event kinematics and angular variables. For the angular variables, the subscript indicates the reference frame. 78

8.2 Expected results for the electron, muon and combined channels. 107

8.3 Observed results for the electron, muon and combined channels. 110

B.1 The standard deviations for the *input-to-hidden* weights, σ_u, *hidden-to-output* weights, σ_v, and all biases, $\sigma_{a,b}$ obtained from a single neural network, JETNET, for the electron channel. 120

B.2 The standard deviations for the *input-to-hidden* weights, σ_u, *hidden-to-output* weights, σ_v, and all biases, $\sigma_{a,b}$ obtained from a single neural network, JETNET, for the muon channel. 121

LIST OF FIGURES

1.1 Properties of fermion particles: quarks and leptons. Spin is given in units of $\hbar$, electric charge is given in units of the absolute value of the electron charge. 3

1.2 Leading order production cross sections for a Standard Model Higgs boson as a function of the Higgs boson mass at 1.96 TeV at the Tevatron $p\bar{p}$ collider. In the cross section calculation the CTEQ6L1 parton distribution function parametrization has been used. 9

1.3 Top quark pair production via the strong interaction: (a) quark-antiquark annihilation; (b) gluon fusion. At the leading order (LO) there are three Feynman diagrams which contribute to the latter process. 10

1.4 Two dominant processes for single top quark production at the Tevatron: (a) s channel process (signal is $W + b + \bar{b}$). (b) t channel process (signal is $W + b + \bar{b} + q$). 13

2.1 Schematic view of the Fermilab accelerator chain. 17

2.2 Schematic view of the Run II DØ detector. 20

2.3 SMT disk and barrel design. 21

2.4 a) Location of the Central Fiber Tracker (CFT). b) Closeup view of axial and stereo layers. 23

2.5 Cross section and layout geometry of CPS and FPS scintillator strips. 24

2.6 Isometric view of the central and two end calorimeters. 25

2.7 Schematic view of two calorimeter cells. 26

2.8 Schematic view showing the calorimeter segmentation pattern. The shading pattern indicates cells for signal readout. The radial lines show the detector pseudo-rapidity intervals. 27

2.9 The DØ muon system. 28

5.1 Reconstruction efficiency of primary vertex. 51

8.1 *First plot*: distributions and p-values for the best modeled variable, p_T(jet3), in the (electron, 1-tag, 4-jet) channel. *Second plot*: distributions and p-values for the worst modeled variable, $\cos(\text{jet1}, \text{alljets})_{\text{alljets}}$. *Third plot*: distributions and p-values for the variable Missing E_T, which has a K-S p-value near the rejection threshold. The 3rd and 4th p-values pertain to the discrepancy measures $D_{\text{minp}} = |C-F|(\text{min } p\text{-value})$ and $D_{\max} = \max|C-F|$, respectively, where C and F, respectively, are the observed and expected counts in the associated local discrepancy, as defined in APPENDIX A. Note that each of the 4 p-values is based on a specific kind of discrepancy between two histograms. Therefore, the 4 p-values need not and indeed, do not, agree exactly. 76

8.2 BNN input variables according to their Rulefit ranking for the electron/1tag channels. 79

8.3 BNN input variables according to their Rulefit ranking for the electron/2tags channels. 80

8.4 BNN input variables according to their Rulefit ranking for the muon/1tag channels. 81

8.5 BNN input variables according to their Rulefit ranking for the muon/2tags channels. 82

8.6 Rows: top = verification plots, bottom = BNN outputs normalized to unity (green: background, blue: signal), and the $S/\sqrt{B}$, $S/\sqrt{S+B}$, and signal efficiency (ϵ_S) versus background efficiency (ϵ_B) curves, for the electron, =1 tag, =2 jet channel. 86

8.7 Rows: top = verification plots, bottom = BNN outputs normalized to unity (green: background, blue: signal), and the $S/\sqrt{B}$, $S/\sqrt{S+B}$, and signal efficiency (ϵ_S) versus background efficiency (ϵ_B) curves, for the electron, =1 tag, =3jet channel. 87

8.8 Rows: top = verification plots, bottom = BNN outputs normalized to unity (green: background, blue: signal), and the $S/\sqrt{B}$, $S/\sqrt{S+B}$, and signal efficiency (ϵ_S) versus background efficiency (ϵ_B) curves, for the electron, =1 tag, =4jet channel. 88

8.9 Rows: top = verification plots, bottom = BNN outputs normalized to unity (green: background, blue: signal), and the $S/\sqrt{B}$, $S/\sqrt{S+B}$, and signal efficiency (ϵ_S) versus background efficiency (ϵ_B) curves, for the electron, =2 tag, =2jet channel. 89

8.10 Rows: top = verification plots, bottom = BNN outputs normalized to unity (green: background, blue: signal), and the $S/\sqrt{B}$, $S/\sqrt{S+B}$, and signal efficiency (ϵ_S) versus background efficiency (ϵ_B) curves, for the electron, =2 tag, =3jet channel. 90

8.11 Rows: top = verification plots, bottom = BNN outputs normalized to unity (green: background, blue: signal), and the $S/\sqrt{B}$, $S/\sqrt{S+B}$, and signal efficiency (ϵ_S) versus background efficiency (ϵ_B) curves, for the electron, =2 tag, =4jet channel. 91

8.12 Rows: top = verification plots, bottom = BNN outputs normalized to unity (green: background, blue: signal), and the $S/\sqrt{B}$, $S/\sqrt{S+B}$, and signal efficiency (ϵ_S) versus background efficiency (ϵ_B) curves, for the muon, =1 tag, =2 jet channel. 92

8.13 Rows: top = verification plots, bottom = BNN outputs normalized to unity (green: background, blue: signal), and the $S/\sqrt{B}$, $S/\sqrt{S+B}$, and signal efficiency (ϵ_S) versus background efficiency (ϵ_B) curves, for the muon, =1 tag, =3jet channel. 93

8.14 Rows: top = verification plots, bottom = BNN outputs normalized to unity (green: background, blue: signal), and the $S/\sqrt{B}$, $S/\sqrt{S+B}$, and signal efficiency (ϵ_S) versus background efficiency (ϵ_B) curves, for the muon, =1 tag, =4jet channel. 94

8.15 Rows: top = verification plots, bottom = BNN outputs normalized to unity (green: background, blue: signal), and the $S/\sqrt{B}$, $S/\sqrt{S+B}$, and signal efficiency (ϵ_S) versus background efficiency (ϵ_B) curves, for the muon, =2 tag, =2jet channel. 95

8.16 Rows: top = verification plots, bottom = BNN outputs normalized to unity (green: background, blue: signal), and the $S/\sqrt{B}$, $S/\sqrt{S+B}$, and signal efficiency (ϵ_S) versus background efficiency (ϵ_B) curves, for the muon, =2 tag, =3jet channel. 96

8.17 Rows: top = verification plots, bottom = BNN outputs normalized to unity (green: background, blue: signal), and the $S/\sqrt{B}$, $S/\sqrt{S+B}$, and signal efficiency (ϵ_S) versus background efficiency (ϵ_B) curves, for the muon, =2 tag, =4jet channel. 97

8.18 Bayesian neural network outputs for the electron channel. [Rows: top =1 tag, bottom =2 tags, columns: left =2 jets, middle =3 jets, right =4 jets.] . . . 98

8.19 Bayesian neural network outputs for the muon channel. [Rows: top =1 tag, bottom =2 tags, columns: left =2 jets, middle =3 jets, right =4 jets.] . . . 99

8.20 BNN outputs from W+jets cross check samples for the electron (left) and muon (right) channels. 100

8.21 BNN outputs from $t\bar{t}$ cross check samples for the electron (left) and muon (right) channels. . . . 101

8.22 Distributions of the measured cross sections (σ_{tb+tqb}), uncertainties ($\Delta\sigma_{tb+tqb}$), and significances ($\sigma_{tb+tqb}/\Delta\sigma_{tb+tqb}$), including all systematic effects, for four different bin and channel combination options overlayed. . . . 103

8.23 The Gaussian distributions (with zero mean and unit standard deviation) sampled for the JES and TRF systematics in the Bayesian analysis. . . . 104

8.24 Distributions of the measured cross sections (peaks of Bayesian posterior density) from the non-SM ensembles. The arrow shows the mean of each distribution. . . . 105

8.25 Ensemble response versus input cross section used to generate the ensemble. The ensemble response is obtained from the mean of the distributions in Fig. 8.24. . . . 106

8.26 Distributions of the expected posterior probability density as a function of the tb+tqb single top cross section with (a) no systematics, and (b) all systematics, and combining the tag and jet multiplicities for different lepton combinations. 108

8.27 Results of the expected measurements in the twelve channels individually and combined. The numbers in parentheses in the right-most column are the expected significances defined as $\sigma_{tb+tqb}/\Delta\sigma_{tb+tqb}$. . . . 109

8.28 Distributions of the posterior probability density as a function of the $tb + tqb$ single top cross section with (a) no systematics, and (b) all systematics, using the measured data and combining the tag and jet multiplicities for different lepton combinations. . . . 111

8.29 Measurements in the twelve channels individually, and combined. . . . 111

8.30 Distribution of cross sections from background-only ensemble with full systematics included. . . . 112

8.31 Distribution of cross sections from SM signal+background ensemble with full systematics included. . . . 113

8.32 The observed BNN output distribution summed over all 12 channels superimposed on the summed signal + background model. The plot on the right is a zoom of the region near a BNN output of 1. . . . 114

A.1 (Electron,1-tag,2-jet)-channel variables ordered according to quality of modeling, quantified by the K-S p-value. The larger the p-value, the less reason one has to reject the hypothesis that the associated variable is poorly modeled, provided of course the kind of discrepancy measured by the K-S statistic is the most appropriate. A different discrepancy measure could yield a different conclusion. 118

B.1 Neural network output distributions for the electron channel as obtained using the JETNET neural network program. The blue histogram is for the sum of all backgrounds, and the red one is for the $tb + tqb$ signal. Rows: top = EqOneTag, bottom = EqTwoTag; columns: left = EqTwoJet, middle = EqThreeJet, right = EqFourJet. 120

B.2 Neural network output distributions for the muon channel as obtained using the single network JETNET algorithm. The blue histogram is for the sum of all backgrounds, and the red one is for the $tb + tqb$ signal. Rows: top = EqOneTag, bottom = EqTwoTag; columns: left = EqTwoJet, middle = EqThreeJet, right = EqFourJet. 121

C.1 The transverse momentum of jet 1 for events with two jets (left column), three jets (center column), and three jets (right column), for electron/1tag (first row), muon/1tag (second row), electron/2tags (third row), and muon/2tags (fourth row). The plot at the bottom of the figure shows the distribution for the electron and muon channels combined, for 2, 3, and 4 jets combined, and for 1 and 2 tags combined. 123

C.2 The transverse momentum of jet 2 for events with two jets (left column), three jets (center column), and three jets (right column), for electron/1tag (first row), muon/1tag (second row), electron/2tags (third row), and muon/2tags (fourth row). The plot at the bottom of the figure shows the distribution for the electron and muon channels combined, for 2, 3, and 4 jets combined, and for 1 and 2 tags combined, i.e., all the channels we have used for this analysis. 124

C.3 The transverse momentum of the electron or muon for events with two jets (left column), three jets (center column), and three jets (right column), for electron/1tag (first row), muon/1tag (second row), electron/2tags (third row), and muon/2tags (fourth row). The plot at the bottom of the figure shows the distribution for the electron and muon channels combined, for 2, 3, and 4 jets combined, and for 1 and 2 tags combined, i.e., all the channels we have used for this analysis. 125

C.4 The missing transverse energy for events with two jets (left column), three jets (center column), and three jets (right column), for electron/1tag (first row), muon/1tag (second row), electron/2tags (third row), and muon/2tags (fourth row). The plot at the bottom of the figure shows the distribution for the electron and muon channels combined, for 2, 3, and 4 jets combined, and for 1 and 2 tags combined, i.e., all the channels we have used for this analysis. 126

C.5 The opening angle $\Delta R(\text{jet1}, \text{jet2})$ for events with two jets (left column), three jets (center column), and three jets (right column), for electron/1tag (first row), muon/1tag (second row), electron/2tags (third row), and muon/2tags (fourth row). The plot at the bottom of the figure shows the distribution for the electron and muon channels combined, for 2, 3, and 4 jets combined, and for 1 and 2 tags combined, i.e., all the channels we have used for this analysis. 127

ABSTRACT

We present results of a search for single top quark production in $p\bar{p}$ collisions using a dataset of approximately 1 fb^{-1} collected with the DØ detector. This analysis considers the muon+jets and electron+jets final states and makes use of Bayesian neural networks to separate the expected signals from backgrounds. The observed excess is associated with a p-value of 0.081%, assuming the background-only hypothesis, which corresponds to an excess over background of 3.2 standard deviations for a Gaussian density. The p-value computed using the SM signal cross section of 2.9 pb is 1.6%, corresponding to an expected significance of 2.2 standard deviations. Assuming the observed excess is due to single top production, we measure a single top quark production cross section of $\sigma(p\bar{p} \to tb+X, tqb+X) = 4.4 \pm 1.5$ pb.

PREFACE

This dissertation is a detailed description of the search for single top quark production using Bayesian neural networks. In Chapter 1, the Standard Model and top quark physics are briefly discussed. Also, the motivation for studying single top quark production is given. Chapter 2 describes the acceleration chain of protons and anti-protons and the DØ detector. The identification of physics objects is explained in Chapter 3. Chapter 4 describes the data set we used. Chapter 5 discusses single top and background modeling, which mostly relies on Monte Carlo simulation. Event selection criteria and expected event numbers after selection are described in Chapter 6. Chapter 7 gives an exposition of the theory and application of Bayesian neural networks. This chapter starts with the basic concepts of (conventional) neural networks and proceeds to explain how to combine Bayesian statistics and neural networks. In addition, Markov Chain Monte Carlo methods, which are used to implement Bayesian neural networks, are described. Chapter 8 explains details of the analysis and result. To validate our result, we performed a statistical test. The interpretation of the test is given in the conclusion.

CHAPTER 1

Introduction

All top quarks discovered and observed at the Tevatron at Fermilab were produced with the anti-top quark through the strong interactions. The Standard Model predicts that top quarks can be produced singly through the weak interactions. This dissertation describes a search for single top quark production and the evidence we found for the existence of this interaction. To discuss the importance of single top quark production, we begin with the fundamental theory, the Standard Model.

1.1 Standard Model

Particle physicists have developed a theory called the Standard Model (SM) [1, 2, 3, 4, 5, 6], which describes all elementary particles and their interactions. The SM is a quantum field theory based on gauge symmetries. It is a very elegant theoretical framework that has successfully passed very precise tests. However, the gravitational interaction is not yet part of this framework.

1.1.1 Quarks and Leptons

There are two types of elementary particles. The first are the basic building blocks of matter. The second are particles that generate interactions. The "matter" particles are fermions and fall into two classes: quarks and leptons. Both quarks and leptons are spin -$\frac{1}{2}$ particles, and therefore obey Fermi-Dirac statistics. As indicated in Fig. 1.1, quarks and leptons are each arranged in three generations, containing particles of similar properties but differing in mass. For each particle there exists an associated anti-particle.

There are six different flavors of quarks, labeled (in order of increasing mass) *up*, *down*, *strange*, *charm*, *bottom*, and *top*. Quarks carry fractional electrical charges of +$\frac{2}{3}$ or -$\frac{1}{3}$.

Leptons spin = 1/2			Quarks spin = 1/2		
Flavor	Mass GeV/c^2	Electric charge	Flavor	Approx. Mass GeV/c^2	Electric charge
ν_e electron neutrino	$<1\times10^{-8}$	0	u up	0.003	2/3
e electron	0.000511	−1	d down	0.006	−1/3
ν_μ muon neutrino	<0.0002	0	c charm	1.3	2/3
μ muon	0.106	−1	s strange	0.1	−1/3
ν_τ tau neutrino	<0.02	0	t top	175	2/3
τ tau	1.7771	−1	b bottom	4.3	−1/3

Figure 1.1: Properties of fermion particles: quarks and leptons. Spin is given in units of $\hbar$, electric charge is given in units of the absolute value of the electron charge.

Quarks are never observed as single particles but form bound states called *hadrons* of either three quarks to form *baryons*, or by pairing a quark with an antiquark to form *mesons*. Protons (made up of two up-quarks and one down-quark) and neutrons (made up of two down-quarks and one up-quark) are the most common examples of baryons. Pions ($\pi^0, \pi^\pm$) and kaons ($K^\pm$, K^0, $\bar{K}^0$) are the most common types of mesons.

There are three different flavors of charged leptons carrying a charge of -1: *electron* (e^-), *muon* (μ^-), and *tau* (τ^-). While electrons exist in all atoms, muons and taus can only be observed in energetic processes like cosmic ray showers, or in high energy particle collisions. There are three neutral leptons, called *neutrinos* (ν), each associated with a charged lepton: ν_e, ν_μ, and ν_τ. Neutrinos interact extraordinarily weakly with matter and their masses are negligibly small.

1.1.2 Interactions

A fundamental interaction (or force) is a mechanism by which particles interact with each other. In nature, there are four fundamental interactions: gravitation, electromagnetism,

the weak interactions, and the strong interactions. Particle interactions are described in terms of exchanges of interaction particles, bosons having spin 1. The photon mediates the electromagnetic interaction, the three weak bosons are the exchanged particles in weak interactions, and the eight gluons are the mediating bosons in the strong interactions.

The Standard Model is a Quantum Field Theory (QFT) in which the interactions manifest certain symmetries. Using the framework of Lagrangian field theory, the theory is required to be invariant under a group of local phase changes (*local gauge invariance*). A local phase depends on the spacetime coordinate. In order to ensure gauge invariance of such a Lagrangian, gauge fields are introduced. These gauge fields lead to spin-1 bosons that are the mediators of the interactions.

Electromagnetic Interactions

Historically, the electromagnetic interactions were the first to be formulated in the framework of a calculable (*renormalizable*) QFT by Tomonaga, Feynman, and Schwinger in the 1940s. *Quantum Electrodynamics* (QED) describes the electromagnetic interactions by requiring gauge invariance under U(1) group transformations. U(1) denotes a group of unitary one-dimensional matrices, describing space-time dependent rotations in a complex plane. The requirement of gauge invariance gives rise to the photon field and the *photons* as the corresponding mediator of the electromagnetic interactions. Because the photon is massless the interaction has infinite range. The photon couples to all particles that carry electrical charge, like quarks and charged leptons. The strength of the interaction is proportional to the magnitude of the dimensionless *fine structure coupling constant*, which is equal to

$$\alpha_{EM} = \frac{e^2}{4\pi} \approx \frac{1}{137}, \tag{1.1}$$

at low momentum transfer.

Weak Interactions

The weak interactions [3] give rise to *beta decays* and associated radioactivity, such as

$$n \rightarrow p + e^{-} + \bar{\nu_e}. \tag{1.2}$$

The range of the interactions is short due to the high mass of the mediating gauge bosons ($W^{\pm}$, Z^0) [7]:

$$m_{W^{\pm}} = 80.425 \pm 0.038 \text{ GeV}, \tag{1.3}$$

$$m_{Z^0} = 91.1876 \pm 0.0021 \text{ GeV}. \tag{1.4}$$

A QFT combining the electromagnetic with the weak interaction was first developed by Glashow, Weinberg, and Salam (*GWS theory*, Nobel Prize in 1979). Later 't Hooft and Veltman were able to prove that the theory is renormalizable. Electroweak theory combines a U(1) group with an SU(2) group, and requires invariance under SU(2)⊗U(1) transformations. SU(n) describes groups of special unitary $n \times n$ matrices. Local gauge invariance under SU(2) group transformations introduces three massless spin-1 gauge fields W^1, W^2, and W^3. Adding the U(1) group introduces another gauge field B^1. The W^3 and B^1 mix quantum mechanically to give rise to the experimentally observed photon (γ) and Z^0:

$$\gamma = W^3 \sin\theta_W + B^1 \cos\theta_W, \tag{1.5}$$

$$Z^0 = W^3 \cos\theta_W - B^1 \sin\theta_W, \tag{1.6}$$

where θ_W is called the *weak mixing angle* or *Weinberg angle*. Unlike QED, the underlying group of the electroweak theory is *non-Abelian* since not all the generators of the group commute with each other.

Up to this point the electroweak theory is very simple and elegant. Yet it is incomplete, since all particles of the theory are massless. Additionally, mass terms cannot be introduced into the Lagrangian describing the system, since this would destroy the local gauge invariance of the Lagrangian. This problem is resolved by the *Higgs mechanism*, which introduces *spontaneous symmetry breaking* of the Higgs scalar field potential, thereby giving mass to the gauge bosons (W and Z) and the quarks and leptons.

Strong Interactions

Quantum Chromo Dynamics (QCD) [8] is the QFT describing the strong interactions. It is based on an SU(3) gauge field, which leads to 8 mediating massless gauge bosons called *gluons*. Quarks carry a new type of "charge" called *color*. Each (anti)quark can carry a (anti)red, (anti)green, or (anti)blue color charge. Gluons carry a combination of a color and anticolor charge. As carriers of the color charge, gluons can couple to each other. This is a consequence of the non-Abelian character of the gauge theory. Quarks and gluons are collectively referred to as *partons*.

One interesting feature of QCD is that the strength of the coupling increases with

decreasing energy scale, i.e. at low energies and long distances the interaction becomes too strong to be treated within the framework of perturbation theory. This leads to *confinement*, which implies that all objects carrying color can never be found as free particles in nature because they are confined in color-neutral hadrons. The quarks that combine into baryons or mesons are referred to as *valence quarks*, and they constantly interact with each other by exchanging gluons. Since gluons can couple to each other, they can emit more gluons that can further split into virtual quark-antiquark pairs called *sea quarks*.

Experimentally, quarks and gluons are manifested as *jets* of color-neutral hadrons. This means that if a single parton emerges from a particle collision, gluons will be radiated which subsequently produce quark-antiquark pairs to form a *parton shower*. Ultimately the partons combine into a jet of hadrons moving in the direction close to that of the original parton. This final step is called *hadronization*.

The strong coupling constant, α_s, can be expressed to leading-log in Q^2 [8] as:

$$\alpha_s(Q^2) = \frac{12\pi}{(11c - 2n_f)\log(\frac{Q^2}{\Lambda^2})}, \tag{1.7}$$

where Q expresses the magnitude of the momentum transferred in the interaction, n_f indicates the number of quark flavors (6 in the SM), and c is the number of quark colors (3 in the SM). Λ is the QCD scale parameter, defined as:

$$\Lambda^2 = \mu_R^2 \exp \frac{-12\pi}{(11c - 2n_f)\alpha_s(\mu_R^2)}. \tag{1.8}$$

The parameter μ_R is an arbitrary renormalization scale introduced to regulate divergences in the perturbative calculation of α_s. Eq. (1.7) shows that the strength of the coupling decreases with increasing momentum transfer Q^2. Therefore, quarks and gluons are said to be *asymptotically free* when probed at high energies. Theoretical work on asymptotic freedom by Gross, Politzer, and Wilczek was rewarded with the 2004 Nobel Prize. On the other hand, as Q^2 approaches Λ, the coupling becomes large and perturbative calculations are no longer possible.

1.1.3 Electro Weak Symmetry Breaking

The electroweak theory is based on the realization that the quantum of light, the photon, and the quanta of β decay, the $W^{\pm}$ bosons, are intimately related. Just as isospin, a symmetry

of strong interactions, identifies the neutron and proton as partners, a new symmetry, weak isospin, identifies an electron and its neutrino as partners.

The electroweak symmetry is far from exact. The W and Z bosons are among the heaviest known elementary particles, while the photon is the lightest, though they are related by this symmetry. Similarly, the neutrino and the electron can hardly be confused, even though they are partners.

How is the electroweak symmetry broken? A magnetic field, for example, applied to an atom breaks its rotational symmetry, or heating up a crystal until it melts breaks the discrete symmetry of the lattice. The theory requires that electroweak symmetry be broken in a similar, though more intricate manner. Without the electroweak symmetry breaking (EWSB), the $W^{\pm}$ and Z, and all the quarks and leptons would be massless. If any progress is to be made in understanding these masses, the source of EWSB must be discovered. It is one of the core questions in high-energy physics.

Mathematically there are several ways to break the electroweak symmetry.

In one approach, a number of scalar particles are introduced. Some of the degrees of freedom provided by three particles are absorbed by the W and Z bosons thus giving them mass. The remaining degrees of freedom appear as new particles; in the simplest realization just one such particle – the Higgs boson.

A second possibility, called supersymmetry, predicts the existence of many new particles, among them a number of scalars like the Higgs boson. While there is no direct evidence for supersymmetry there is strong theoretical motivation for it. There is also some supporting circumstantial evidence from extrapolating the electroweak and strong couplings to high energies, where the three couplings coalesce – if supersymmetry effects are included – as they should if there is a grand unification of these forces at high energy.

A third possibility, referred to as strongly coupled EWSB, introduces no new particles but requires that their role is played by new features of the strong interactions. The corresponding theory is commonly referred to as technicolor.

1.1.4 The Standard Model Higgs

The simplest form of EWSB is realized with a doublet of complex scalar fields that introduce four new degrees of freedom of which a single neutral scalar particle, the Higgs boson, remains after symmetry breaking.

At hadron colliders, Higgs bosons can be produced via four different production mechanisms:

- gluon fusion, $gg \rightarrow H$, which is mediated at lowest order by a heavy quark loop;
- vector boson fusion (VBF), $qq \rightarrow qqH$;
- associated production of a Higgs boson with weak gauge bosons, $qq \rightarrow W/ZH$ (Higgs Strahlung, Drell-Yan like production);
- associated Higgs boson production with heavy quarks, $gg, qq \rightarrow ttH$, $gg, qq \rightarrow bbH$ (and $gb \rightarrow bH$).

The lowest order production cross sections for the four different processes are shown in Fig. 1.2 for the Tevatron collider as a function of the Higgs boson mass. The dominant production mode is the gluon-fusion process. In the low mass region it amounts at leading order to about 20% of the gluon-fusion cross section, whereas it reaches the same level for masses around 800 GeV/c^2. At the Tevatron $p\bar{p}$ collider, the contribution of the associated W/ZH production mode is also important and Higgs boson searches exploit this production mode.

The most relevant decays of the SM Higgs boson are summarized in Ref. [9, 10]. For masses below about 130 GeV/c^2, decays to fermion pairs dominate, of which the decay $H \rightarrow b\bar{b}$ has the largest branching ratio. Decays to $\tau^+\tau^-$, $c\bar{c}$ and gluon pairs (via loops) contribute less than 10%. For such low masses the decay width is less than 10 MeV/c^2. For larger masses the W^+W^- and ZZ final states dominate and the decay width rises rapidly, reaching about 1 GeV at $m_h = 200$ GeV/c^2 and even 100 GeV/c^2 at $m_h = 500$ GeV/c^2.

The direct search for the Higgs boson at the e^+e^- collider LEP has led to a lower bound on its mass of 114.4 GeV/c^2 [11]. Indirectly, high precision electroweak data constrain the mass of the Higgs boson via their sensitivity to loop corrections. Assuming the overall validity of the Standard Model, a global fit [12] to all electroweak data leads to $m_h = 114^{+69}_{-45}$ GeV/c^2. On the basis of the present theoretical knowledge, the Higgs sector in the Standard Model remains largely unconstrained. While there is no direct prediction for the mass of the Higgs boson, an upper limit of ≈ 1 TeV/c^2 can be inferred from unitarity arguments [13].

Further constraints can be derived under the assumption that the Standard Model is valid only up to a cutoff energy scale Λ, beyond which new physics becomes relevant. Requiring

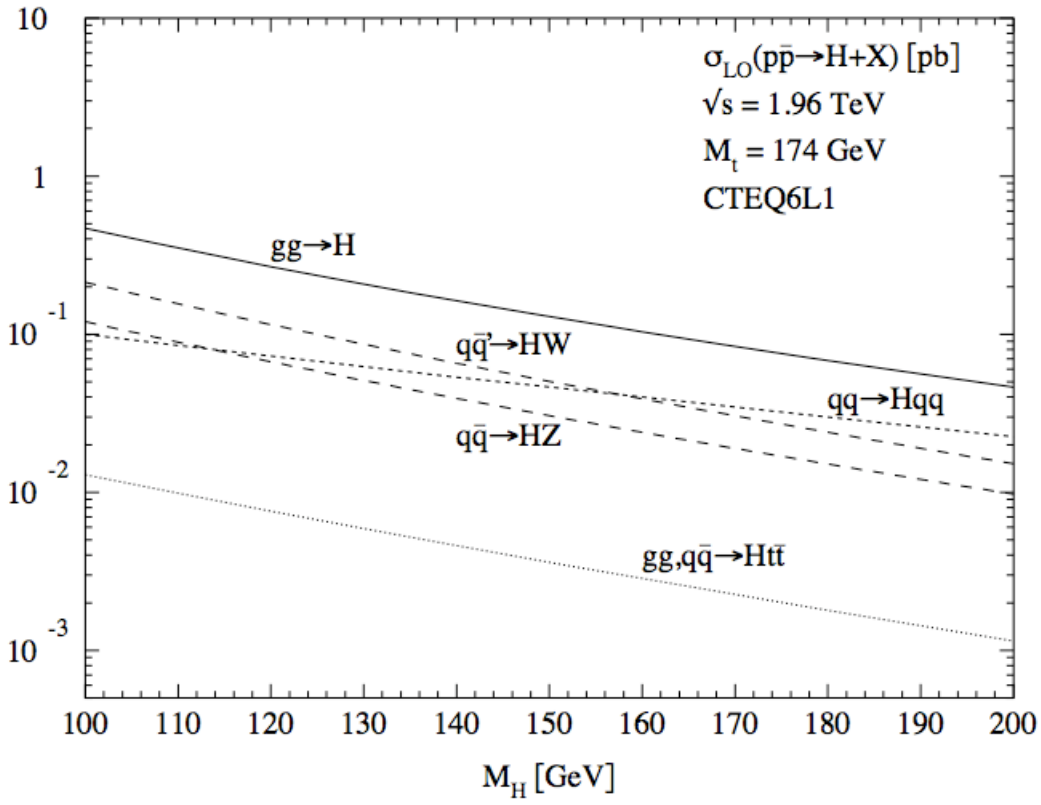


Figure 1.2: Leading order production cross sections for a Standard Model Higgs boson as a function of the Higgs boson mass at 1.96 TeV at the Tevatron $p\bar{p}$ collider. In the cross section calculation the CTEQ6L1 parton distribution function parametrization has been used.

that the electroweak vacuum be stable and that the Standard Model remains perturbative allows one to set upper and lower bounds on the Higgs boson mass [14, 15]. For a cutoff scale of the order of the Planck mass, the Higgs boson mass is required to be in the range $130 < m_h < 190$ GeV/c^2. If new physics appears at lower mass scales, the bound becomes weaker, e.g., for $\Lambda = 1$TeV/c^2 the Higgs boson mass is constrained to be in the range $50 < m_h < 800$ GeV/c^2.

Upper bounds, obtained by the Tevatron experiments CDF and DØ for the cross sections of event topologies motivated by Higgs boson production in the SM are shown in Ref. [16].

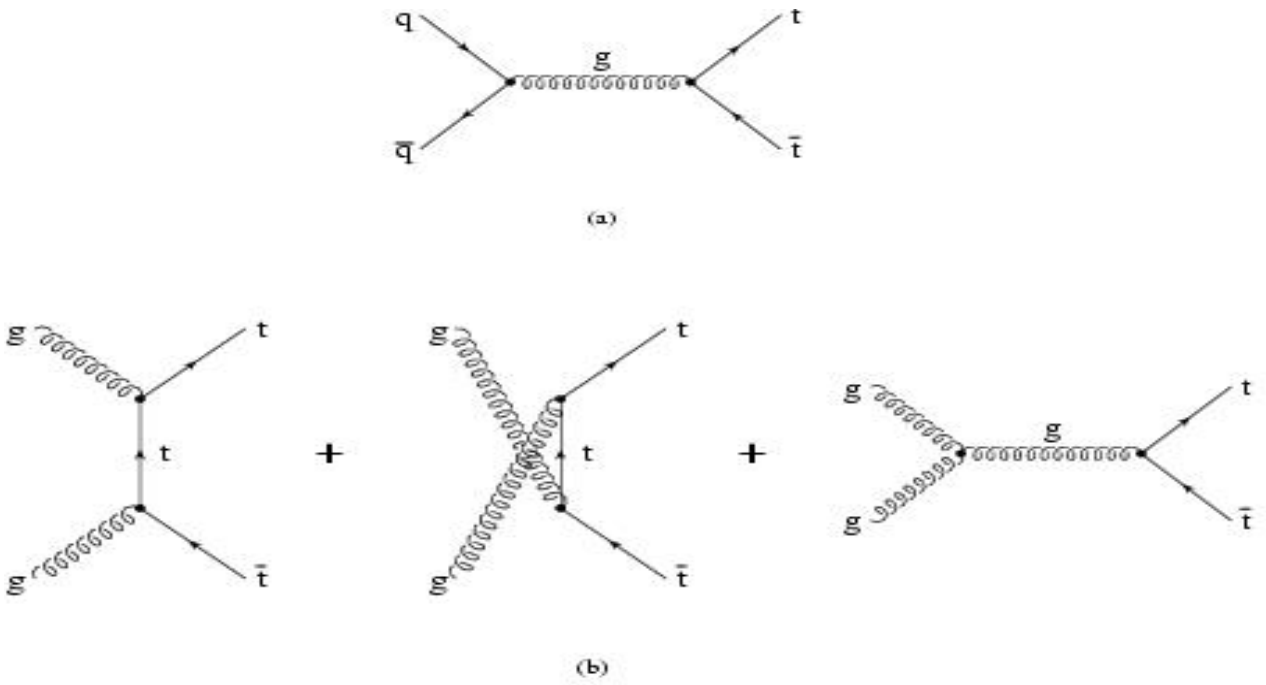


Figure 1.3: Top quark pair production via the strong interaction: (a) quark-antiquark annihilation; (b) gluon fusion. At the leading order (LO) there are three Feynman diagrams which contribute to the latter process.

1.2 The Top Quark

In 1995, the discovery of the top quark was announced by DØ and CDF experiments at Fermilab [17, 18]. After nearly 20 years since the discovery of the b quark, physicists observed an isospin partner of a third generation down-type quark. The top quark indicates most of the properties of an up-type quark, interacting via the weak and strong forces, having a charge of +2/3e, and being a spin 1/2. On the other hand, it exhibits several unique properties. It has a very short life time and a very large mass. Its life time is about 0.4×10^{-24} s, which is shorter than the characteristic hadronization time scale $\sim 3.0\times10^{-24}$ s. Therefore, top bound states do not have time to form, and the top quark can be studied as a free particle [19]. The mass of the top quark is measured to be 171.4 ± 2.1 GeV [20]. It is much heavier than the next to heaviest quark, the b quark, which has a mass of about 5 GeV.

The top quark is produced in $p\bar{p}$ collisions mostly via the strong interactions along with its antiparticle ($t\bar{t}$ pair production). At $\sqrt{s} = 1.8$ TeV (Run I) top pairs were produced 90%

of the time via the quark-antiquark annihilation process ($q\bar{q}$), and the remainder of the time via gluon-gluon process (gg). In Run II, the fraction of the $q\bar{q}$ process decreases to 85%. Within the SM the top quark decays into $W + b \sim 100\%$ of the time. The W decays with the following branching ratio (BR):

W^+	$e^+\nu$	$\mu^+\nu$	$\tau^+\nu$	$u\bar{d}$	$c\bar{s}$
BR	1/9	1/9	1/9	3/9	3/9

For $t\bar{t}$ pair production the event topologies are:

- Dilepton channel ($l\nu l\nu b\bar{b}$): events for which both W's decay into e or μ. This is expected to occur with a branching ratio of 4/81, i.e., $\sim 5\%$ of the final states.
- Lepton + jets channel ($l\nu q\bar{q}b\bar{b}$): events in which one W decays into e or μ, the other into a quark pair. This occurs with a branching ratio of 24/81, i.e. in $\sim 30\%$ of the events.
- All-jets channel($q\bar{q}q\bar{q}b\bar{b}$): events in which both W's decay into quark pairs. This occurs with a branching ratio of 36/81, i.e. in $\sim 44\%$ of the events.

An event with a τ and another lepton is typically not considered as part of the dilepton channel since τ decays further. The dilepton channel comprises $\mu\mu$+jets, μe+jets, or ee+jets, and two neutrinos in each event. The main difficulty of this channel lies in reconstructing two W bosons because there is only one missing transverse energy measured in the detector. The all jet channel consists of two b jets from two top quark decays and four or more light quark jets from the W boson decays. To identify the jets from each decayed top quark correctly, sophisticated statistical methods are demanded. The lepton+jets channel is relatively easier than other channels in $t\bar{t}$ measurements since it has a clean W boson from the lepton and missing transverse energy and fewer jet combinatorics. DØ and CDF observed about one hundred $t\bar{t}$ events in Run I. The measurement of the cross section and mass by DØ and CDF are:

$$\sigma = 5.9 \pm 1.7 \text{ pb}, \; m_t = 172.1 \pm 7.1 \text{ GeV (DØ)} \; [21],$$
$$\sigma = 6.5^{+1.7}_{-1.4} \text{ pb}, \; m_t = 176.0 \pm 6.5 \text{ GeV (CDF)} \; [22].$$

The combined mass is:

$$m_t = 174.3 \pm 5.1 \text{ GeV [23, 24].}$$

In RunII, the DØ and CDF experiments have made a large number of precision top quark measurements and the results are:

$$\sigma = 6.4^{+1.3}_{-1.2}(stat) \pm 0.7(sys) \pm 0.4(lumi) \text{ pb (DØ) [25],}$$
$$\sigma = 7.3 \pm 0.5(stat) \pm 0.6(sys) \pm 0.4(lumi) \text{ pb (CDF) [26],}$$
$$m_t = 171.4 \pm 2.1 \text{ GeV [20].}$$

1.3 Single Top Quark

The Standard Model predicts that top quarks can be created in pairs via the strong force, or singly via the electroweak interaction. The second production mode is referred to as single top quark production and takes place mainly through the s or t channel exchange of a W boson. Another single top production, the tW mode occurs when a b quark radiates a W. However, this process has a negligible cross section at the Tevatron and we will not discuss it further. Figure 1.2 illustrates the leading order Feynman diagrams of s and t channel productions. Each process is described below.

- s channel production: This process involves quark anti-quark annihilation with an off-shell W^* boson and produces a top quark and a $\bar{b}$ quark. The s channel is referred to as the tb mode which includes both $t\bar{b}$, $\bar{t}b$ and only contains the 2→2 process at leading order.

- t channel production: This process includes a 2→2 part with a b quark from the proton sea in the initial state, and a dominant 2→3 part, where an extra $b\bar{b}$ anti-quark appears in the final state explicitly. It is also called W-gluon fusion for the 2→3 process and allows a study of the charge-current weak interaction of the top quark. We refer to the t channel process as tqb, which includes $tq\bar{b}$, $\bar{t}\bar{q}b$, tq, and $\bar{t}\bar{q}$.

1.3.1 Motivation to Study Single Top Production at Hadron Colliders

Studying single top production is important because it can be used to investigate top quark properties that are not accessible through $t\bar{t}$ measurements alone. Several reasons to study single top production are described below.

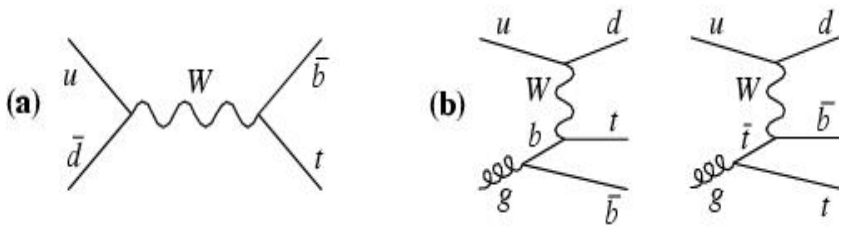


Figure 1.4: Two dominant processes for single top quark production at the Tevatron: (a) s channel process (signal is $W + b + \bar{b}$). (b) t channel process (signal is $W + b + \bar{b} + q$).

- The study of single top quark production provides the only window into measuring the Cabibbo-Kobyashi-Maskawa (CKM) matrix element [27] V_{tb}, which is closely tied to the number of quark generations.
- Measuring the spin polarization of single top quarks can be used to test the V-A structure of the top weak charged current interaction.
- Single top events represent an irreducible background to several searches for SM or non-SM signals, for example, Higgs boson searches.
- The presence of various new SM and non-SM phenomena may be inferred by observing deviations from the predicted rate of the single top signal.

Measurement of V_{tb}

Top quark decays into a W boson and a down-type quark. The flavor of the down-type quark is determined by the CKM matrix.

$$\begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix} \begin{bmatrix} d \\ s \\ b \end{bmatrix} = \begin{bmatrix} d' \\ s' \\ b' \end{bmatrix}.$$

To measure V_{tb} in the Standard Model, one strong assumption is required: there are only three quark generations in nature. Then the unitarity of the CKM matrix implies

$$|V_{ub}|^2 + |V_{cb}|^2 + |V_{tb}|^2 = 1,$$

and V_{tb} is restricted to

$$0.9990 \leq |V_{tb}| \leq 0.9993.$$

With this assumption, the ratio of top decays into b quarks and of top decays into all quarks has been measured at the Tevatron [28]

$$\frac{B(t \to Wb)}{B(t \to Wq)} = \frac{|V_{tb}|^2}{|V_{td}|^2 + |V_{ts}|^2 + |V_{tb}|^2} = 0.94^{+0.31}_{-0.24} \; .$$

If we do not assume three generations, then V_{tb} will be virtually unconstrained:

$$0.08 \leq |V_{tb}| \leq 0.9993, \; |V_{tb}| \gg |V_{ts}| \, , |V_{td}| \, .$$

The interpretation of the direct measurement of V_{tb} depends on the ratio σ_m/σ_t. Here, σ_m is the measured cross section for single top and σ_t is the theoretical cross section for single top. If the ratio differs from one, indicating a cross section different from the Standard Model expectations, then the CKM matrix element V_{tb} must deviate from unity and the likely explanation might be observation of new physics. Therefore, one of the purposes of a direct measurement of V_{tb} is to search for new physics.

Measurement of Spin Polarization

The single top quark is produced through a left-handed interaction and therefore it is expected to be highly polarized. Since no hadronization occurs, spin correlations survive in the final decay products. Hence, single top quark production offers an opportunity to observe the polarization and to test the remarkable SM predictions. It has been shown [29] that the top quark spin in each event follows the direction of the down-type quark momentum in the top quark rest frame. This is the direction of the initial $\bar{d}$-quark for the s channel, and mostly the direction of the final d-quark for t channel single top production. It has been pointed out [30] that the above result follows directly from the properties of the polarized top decays when single top production is considered as top quark decay going "backwards in time." The decay differential width of a polarized top quark to a bottom quark and two leptons or two light quarks is given by a very simple formula in the Standard Model

$$\frac{1}{\Gamma}\frac{d\Gamma}{d\cos\theta^*_{fp}} \;\; = \;\; \frac{1}{2}(1 + K_f \cos\theta^*_{fp}),$$

where θ^*_{fp} is the angle between the momentum direction of one of the final fermions f in the top rest frame and the direction of the top quark polarization vector. The coefficients K_f are equal to 1 for the down-type fermions l^+, d and s quarks, and to -0.31 for the up-type fermions ν_l, u and c quarks [31]. This means that the down-type fermions are the best

top quark spin analyzers. The next-to-leading order (NLO) corrections do not change this property significantly. NLO corrections to the lepton factor K_l are very small, $-0.0015\alpha_s$ [32] and to the quark factor $K_{d,s}$ they are about -6% [33].

From consideration of single top production as a decay going "backwards in time", one can easily conclude that the best variable to observe maximal top spin correlations between single top production and subsequent decay is the angle between the aforementioned d-quark direction in the production processes and the charged lepton (or d, s-quark) direction from the top decay in the top rest frame.

Finally, measurements of the charged-current couplings of the top quark may probe any nonstandard structure of the couplings and therefore provide hints of new physics. Especially any deviation in the $(V–A)$ structure of the Wtb coupling would lead to a violation of the spin correlation properties [34, 35].

CHAPTER 2

Accelerator and Detector

The Tevatron collider, located at the Fermi National Accelerator Laboratory (Fermilab), provides proton and antiproton collisions with a center-of-mass energy of 1.96 TeV. It is the world's highest energy accelerator and the only place to study top quark physics directly. Two general purpose detectors (DØ and CDF) are placed at intersecting regions in the Tevatron where collisions take place. These detectors are designed to identify particles produced from the collisions using 36×36 proton and antiproton bunches, and 396 ns between bunch crossings. This analysis uses data collected by the DØ detector, which consists of a central tracking system, a calorimeter, and a muon system. In the following sections, we describe the Tevatron accelerator and the DØ detector.

2.1 Tevatron Accelerator

To achieve a center-of-mass collision energy of 1.96 TeV, a series of steps to accelerate beams of protons and antiprotons is required [36, 37, 38]. The Tevatron complex has five main accelerators and storage rings linked together.

- The Pre-Accelerator
- The Linac
- The Booster
- The Main Injector
- The Antiproton Source
- The Tevatron

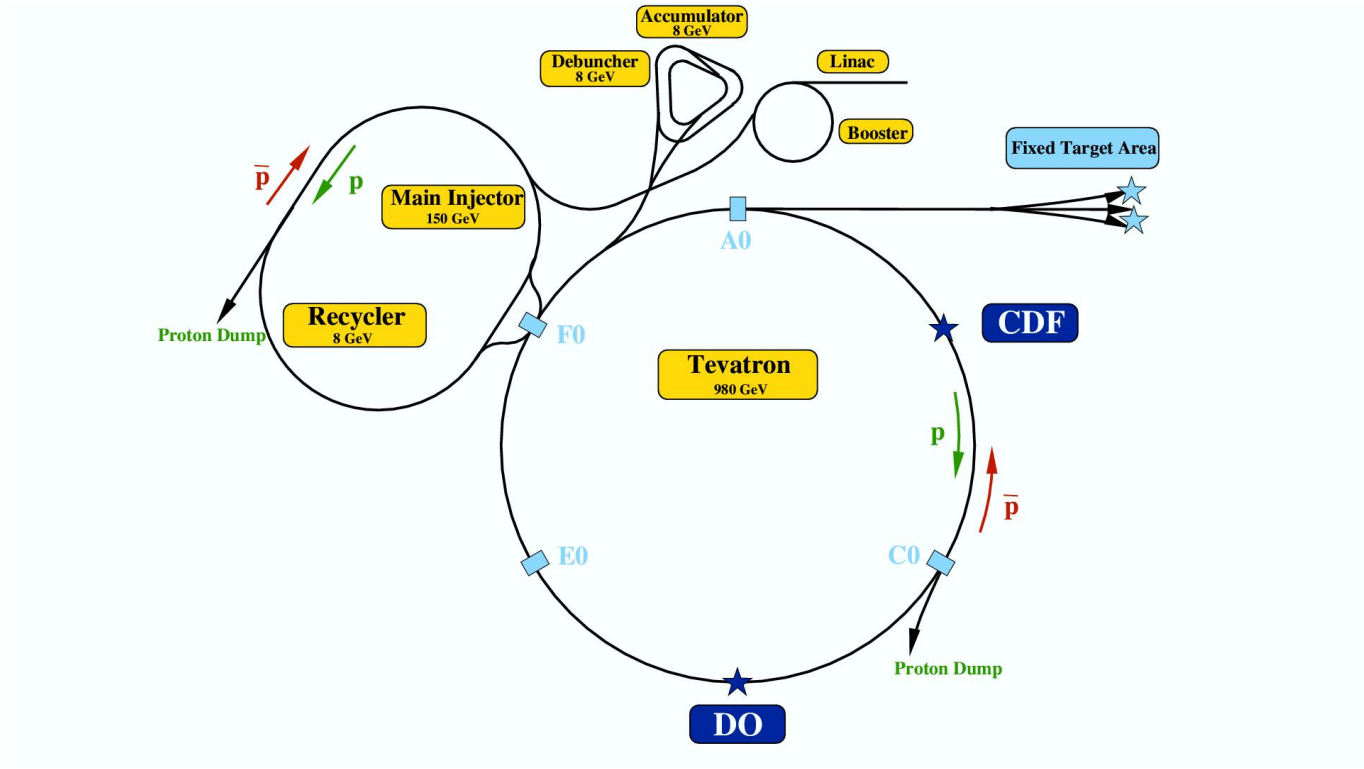


Figure 2.1: Schematic view of the Fermilab accelerator chain.

The Pre-Accelerator

The acceleration chain begins with the proton source in the pre-accelerator. Negatively charged hydrogen atoms are brought up to an energy of 750 keV using a Cockcroft-Walton accelerator. To accelerate charged particles, one of the simplest ways is to place the particles within a constant electric field, generated by two electrodes with different potential. The Cockcroft-Walton accelerator produces a 750 kV potential difference by charging capacitors in parallel from an AC voltage source and discharging them in series and increases the energy of H^- ions from 18 keV to 750 keV.

The Linac

Since very large energy differences lead to sparking between the electrodes, electrostatic accelerators can only be used for a limited energy range. To accelerate H^- ions from the pre-accelerator further, a linear accelerator (Linac) is used. The Linac uses alternating

radiofrequency (RF) fields. A charged particle exposed to an RF field will experience no net acceleration because the accelerating and decelerating phases of the field will cancel one another. However, in the Linac, H^- ions are shielded by RF cavities during the decelerating phases of the RF field so the acceleration can be made non zero. After passing through the Linac, bunches of 400 MeV H^- ions are transferred into the Booster.

The Booster

A circular synchrotron accelerator, called the Booster, is used for further acceleration. The Booster is the first synchrotron in the chain of accelerators and uses dipole electromagnets to steer charged particles into a circular path of fixed radius. In the Booster the H^- ions are stripped of their electrons leaving protons only. The protons are accelerated up to 8 GeV and pass to the Main Injector.

The Main Injector

The main injector accelerates protons coming from the booster and antiprotons coming from the antiproton source from 8 GeV to 150 GeV. It also delivers 120 GeV protons to the antiproton source.

The Antiproton Source

120 GeV protons are made to collide with a nickel target to produce 8 GeV antiprotons. The target consists of 10 cm diameter nickel disks separated by copper cooling disks. In approximately one out of every 100,000 collisions, an antiproton is produced. The antiprotons are sent to the debuncher ring, and a process called stochastic cooling is used to reduce the momentum spread of the particles. They are temporarily stored in the accumulator and inserted into the main injector.

The Tevatron

The Tevatron is the final stage in the chain of proton and antiproton acceleration. It uses superconducting niobium/titanium magnets to provide magnetic fields of up to 4 Tesla, allowing for acceleration to a center-of-mass collision energy of 1.96 TeV. Bunches from the main injector are coalesced into a single bunch and passed into the Tevatron. In each store, 36 bunches of both protons and antiprotons are injected and travel in opposite directions

while sharing the same beam pipe. Collisions occur at DØ and CDF regions by means of focusing magnets, which tighten the diameter of the proton and antiproton beams to 40 μm. The proton and antiproton beams cross one another at each detector every 396 ns [39].

2.2 DØ Detector

The DØ detector [40, 41] consists of three major components: a central tracking system that provides precision vertex and momentum measurements, a liquid-argon/uranium calorimeter that measures particle energies, and a muon system.

2.2.1 Coordinate System

In the DØ detector, the coordinate system is right-handed. The direction of the proton beam is defined as the positive z-axis, and the positive y-axis is pointing upwards. In most cases, cylindrical coordinates are used. The polar angle $\theta = 0$ coincides with the positive z-axis, and the azimuthal angle $\phi = 0$ with the positive x-axis, pointing away from the center of the Tevatron. The polar angle is more commonly described by the pseudorapidity η:

$$\eta = -\ln\left[\tan\frac{\theta}{2}\right] \tag{2.1}$$

The pseudorapidity approximates the true rapidity in the limit of $m \ll E$,

$$y = \frac{1}{2}\cdot\ln\left[\frac{E+p_z}{E-p_z}\right]. \tag{2.2}$$

Zero pseudorapidity corresponds to particles emitted at 90° relative to the beamline, and high values of $|\eta|$ correspond to forward or backward going particles.

2.2.2 Tracking System

The tracking system consists of two subsystems, the Silicon Microstrip Tracker (SMT) [42], surrounded by the Central Fiber Tracker (CFT) [43], contained within a 2T superconducting solenoid. These systems are new for Run II. With this tracking system, DØ can measure momenta of charged particles more precisely, the primary collision vertex with a resolution of 35 μm along the z axis, and the trajectory of particles with a large range of pseudorapidity ($|\eta| < 3$).

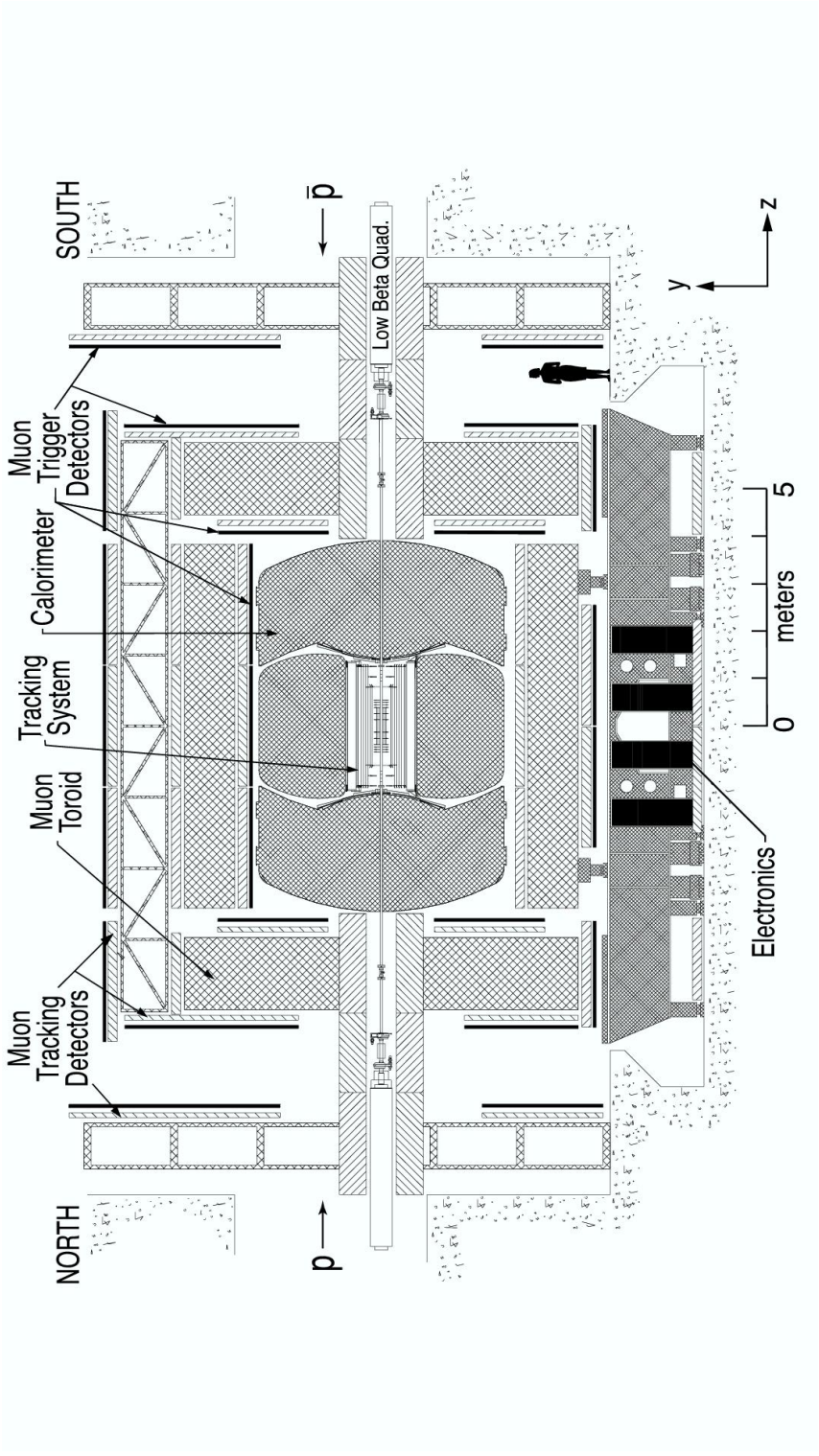


Figure 2.2: Schematic view of the Run II DØ detector.

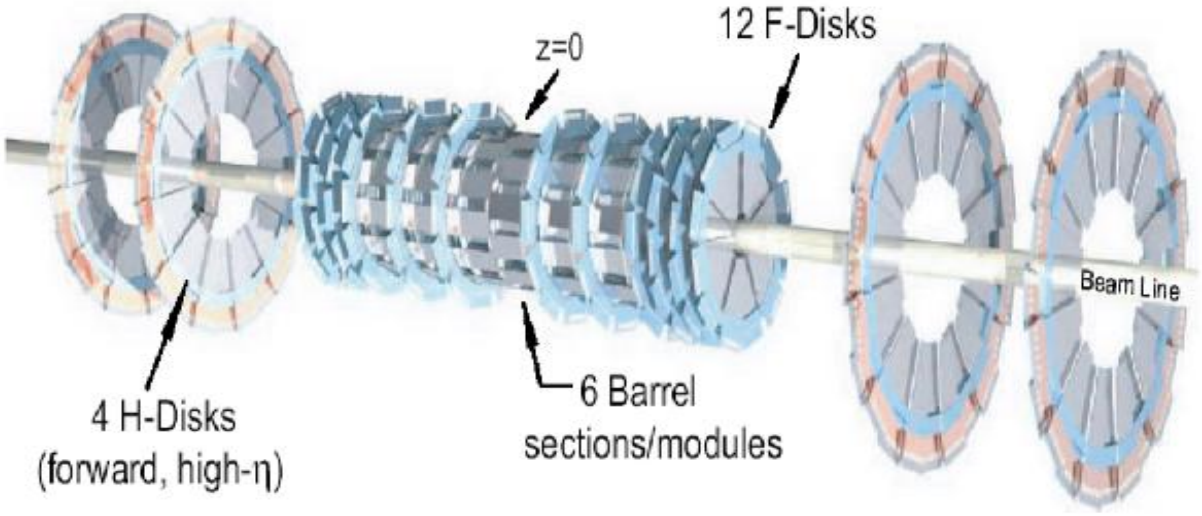


Figure 2.3: SMT disk and barrel design.

Silicon Microstrip Tracker (SMT)

The silicon microstrip tracker (SMT) is designed to provide high resolution tracking. The basic detection unit is an n-type silicon wafer with p-type or n^+-type parallel microstrips across the surface. When charged particles pass through n-type wafers, they ionize the silicon and produce current flows, measured by the microstrips. These are known as "hits" and provide the position of the ionization in one dimension. The thickness of the wafer is 300μm and the pitch between adjacent strips is 50μm.

The SMT has a barrel system consisting of four layers of silicon wafers measuring the $r-\phi$ coordinate and disk structures of twelve wedge detectors measuring the $r-z$ as well as $r-\phi$ coordinates. Six barrels and disks, called "F-disk", are arranged alternately, capped at the ends by additional groups of 6 F-disk detectors and 4 large diameter H-disks (Fig. 2.3).

All silicon sensors are doublesided except for the sensors of layer 1 and 3 of the outermost barrels. In layers 2 and 4, microstrips have a stereo angle of 2° and layers 1 and 3 have microstrips oriented at 90° stereo. In F-disks, silicon microstrips have ±15° orientation.

H-disks are made of 24 wedges and each wedge consists of two single sided wedges glued back-to-back, with an effective 15° stereo angle.

Central Fiber Tracker (CFT)

The second nearest detector to the $p\bar{p}$ collision point outside the SMT is the central fiber tracker (CFT) which enables track reconstruction and momentum measurement of all charged particles at $|\eta| < 2.0$. It uses scintillating fibers to detect the passage of charged particles. The fibers emit light as particles travel through them. This light reaches the visible light photon counters (VLPCs) through waveguides. A VLPC is a silicon avalanche photodetector with an operating temperature of 9 K. A VLPC is an ideal readout device because it provides high gain, fast reset time, and excellent quantum efficiency.

The scintillating fibers are arranged in eight concentric cylinders ranging in radius from 20 cm to 52 cm. The diameter of a fiber is 835 μm and the two inner cylinders are 1.66 m long, while the remaining cylinders are 2.52 m long. Each cylinder has two ribbons containing two layers of 128 fibers. The second layer is offset from the first by half a diameter, as in (Fig. 2.4). In the cylinder, one ribbon is aligned along the z axis, and another ribbon is aligned at a stereo angle of ±3° (Odd numbered cylinders use +3° and even numbered cylinders use -3°).

2.2.3 Preshower Detector

The preshower detector [44, 45] is designed to enhance electron identification and the calorimetric measurement by sampling the shower multiplicity after traversing the material in the solenoid. It is used as an extension to the calorimeter and as a final layer to the central tracking. The central preshower detector (CPS) with a coverage of $|\eta| < 1.3$ is located in the 51 mm gap between the solenoid coil and the central cryostat at a radius of 72 cm. The CPS consists of triangular scintillator strips with one axial layer and two stereo layers. (Fig. 2.5) shows a cross sectional view of the geometry. Between the CPS and the solenoid a thin lead radiator (1 radiation length) encased by two stainless plates is tapered.

The forward preshower detector (FPS) covers a range of $1.4 < |\eta| < 2.5$ and is located on the inner face of each of the end calorimeter cryostat. The FPS is made of two layers of double layered scintillator strips, separated by a 11 mm thick lead-steel plate. The first layer is called the minimum ionizing particle (MIP) layer and the second layer, behind the

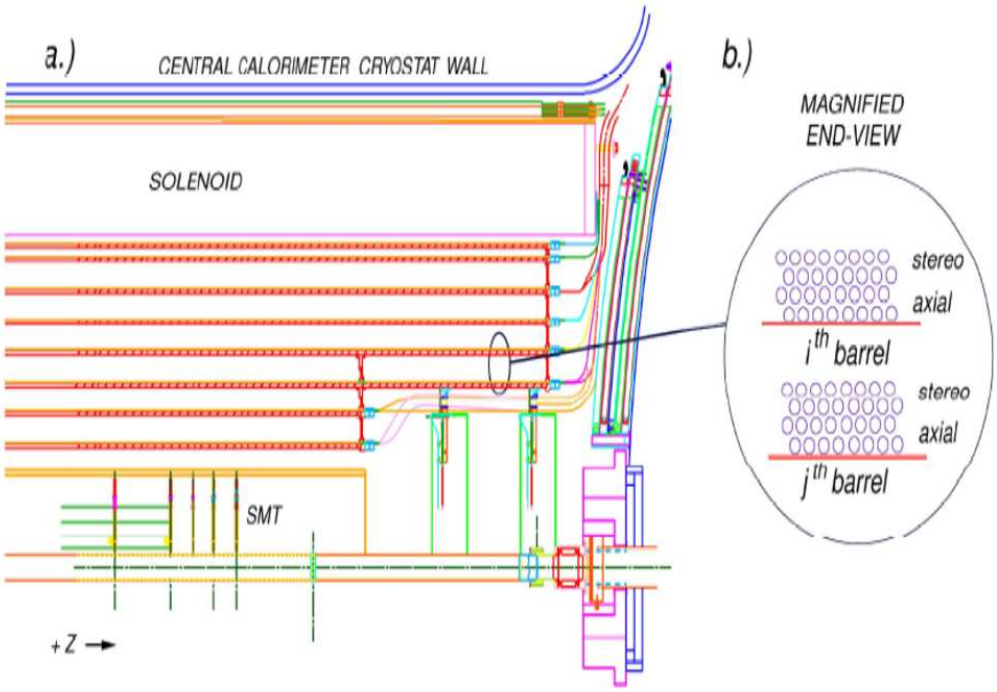


Figure 2.4: a) Location of the Central Fiber Tracker (CFT). b) Closeup view of axial and stereo layers.

lead plate, is known as the shower layer. Heavy charged particles will register a hit only in the MIP layer. However, electrons will leave a similar hit in the MIP layer and will begin to shower in the shower layer, creating a cluster of energy. Photons leave no signal in the MIP layer, but produce a shower signal in the shower layer.

2.2.4 The Calorimeter

The DØ calorimeter measures the energies of electron, photon and jets. It cannot detect neutrinos and identifies only a minimally-ionizing particle signature for muons. The calorimeter consists of three units, a central calorimeter (CC) which covers up to $\eta \approx 1.0$, and a pair of end cap calorimeters (EC), which provide energy measurements in the region $1.4 < \eta < 4.0$.

Each calorimeter region is segmented into many cells containing an absorber medium, an active medium, and a copper readout pad (Fig. 2.7). The surface of the pad is held

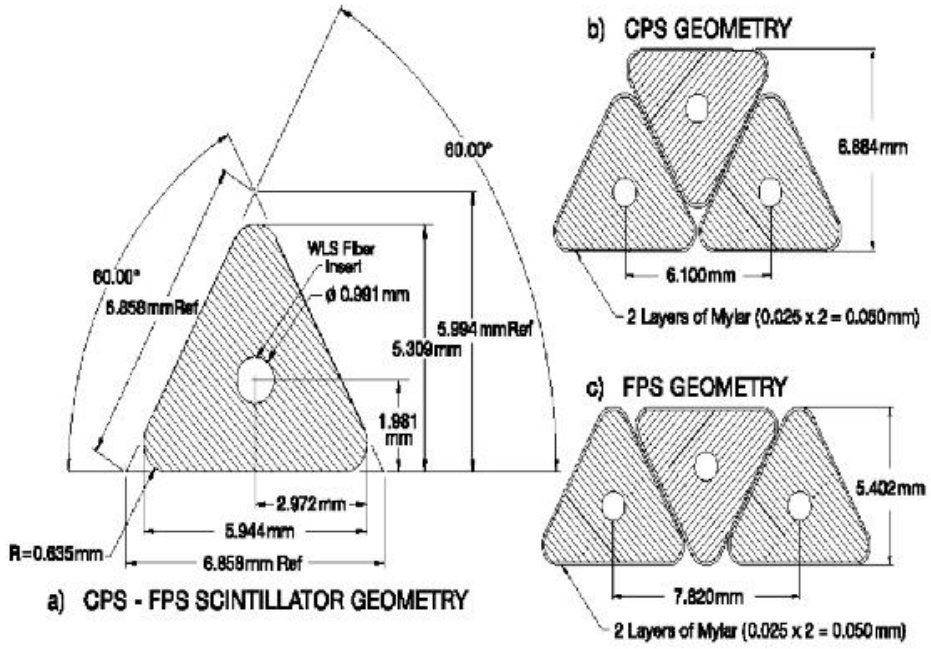


Figure 2.5: Cross section and layout geometry of CPS and FPS scintillator strips.

at a positive voltage to collect the ionization; particles entering a cell are slowed down by the absorber and ionize the active medium. The readout pad measures a current from the ionizing particles which is proportional to the energy deposited in that cell. Liquid argon is used as the active medium because of the good radiation hardness, the flexibility offered in segmenting the calorimeter into transverse and longitudinal cells, and the relatively low cost for readout electronics. However, the use of liquid argon requires the calorimeters be housed in a massive containment vessel (cryostat), which leads to regions of uninstrumented material, and inaccessibility of modules during operation.

In the CC, there are three types of modules: an electromagnetic section (EM), a fine hadronic section (FH), and a coarse hadronic section (CH). The EM section is designed to measure the energy of electromagnetic particles and photons precisely. It works on the principle that a collision between an absorber, which is a relatively thin (several mm) uranium plate, and an incident particle will induce a shower of particles. The EM calorimeter is further divided into four layers. The first two layers measure the longitudinal shower development near the beginning of the shower. The third layer is placed where the shower is expected to

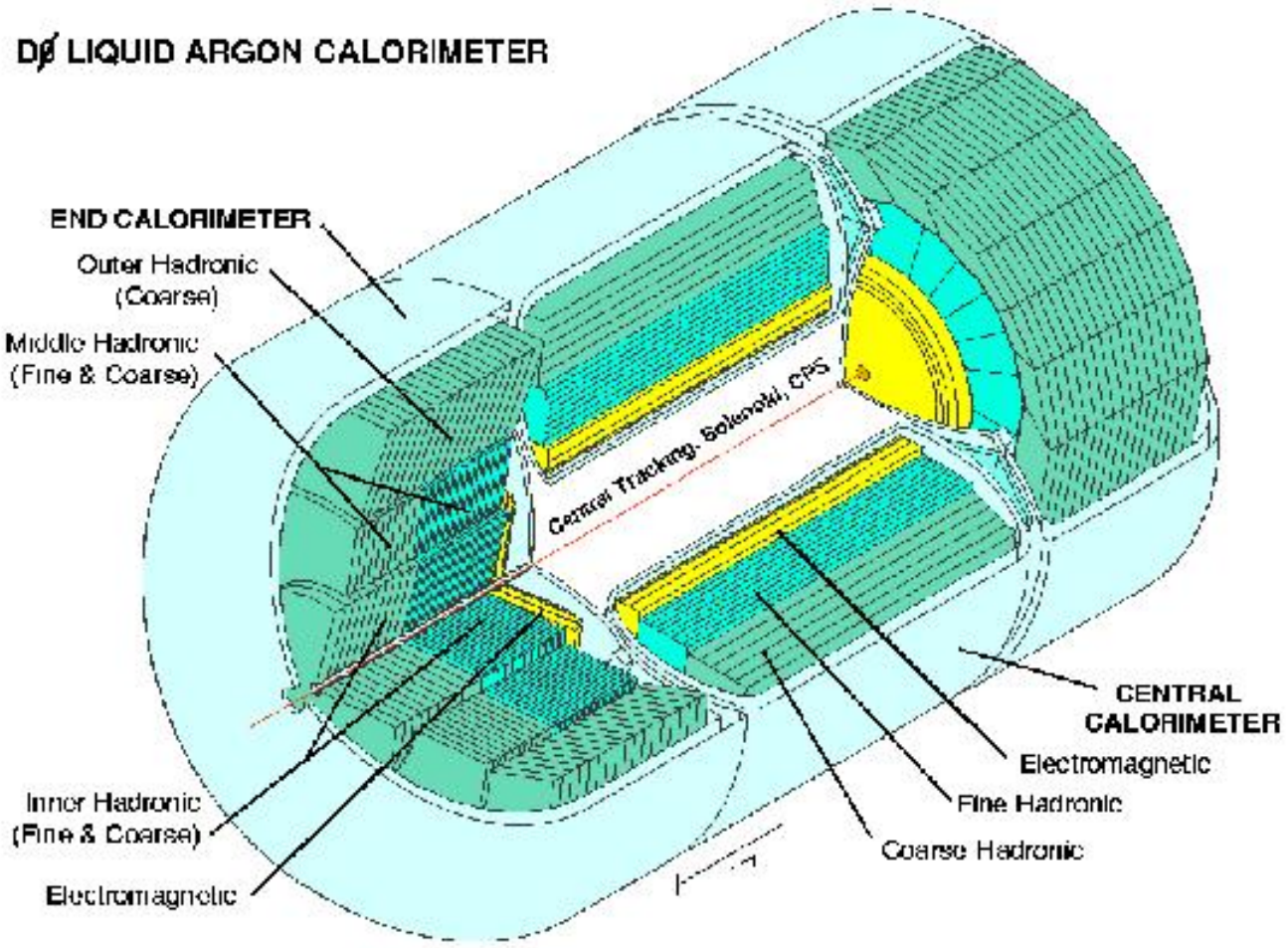


Figure 2.6: Isometric view of the central and two end calorimeters.

reach its maximum and has twice the usual segmentation in both η and ϕ to provide more precise location of EM showers. The fourth layer completes the EM coverage at 20 radiation lengths.

The hadronic calorimeters (FH and CH) work similarly. Showers are produced by the collisions of hadrons with an absorber. The FH section has a thicker ($\approx$ 6 mm) uranium plate absorber and consists of three layers used to measure the further penetrating hadronic showers. The CH section has a thick ($\approx$ 5 cm) copper or stainless steel plate absorber and consists of a single thick layer used to effectively contain the remaining energy in the particle

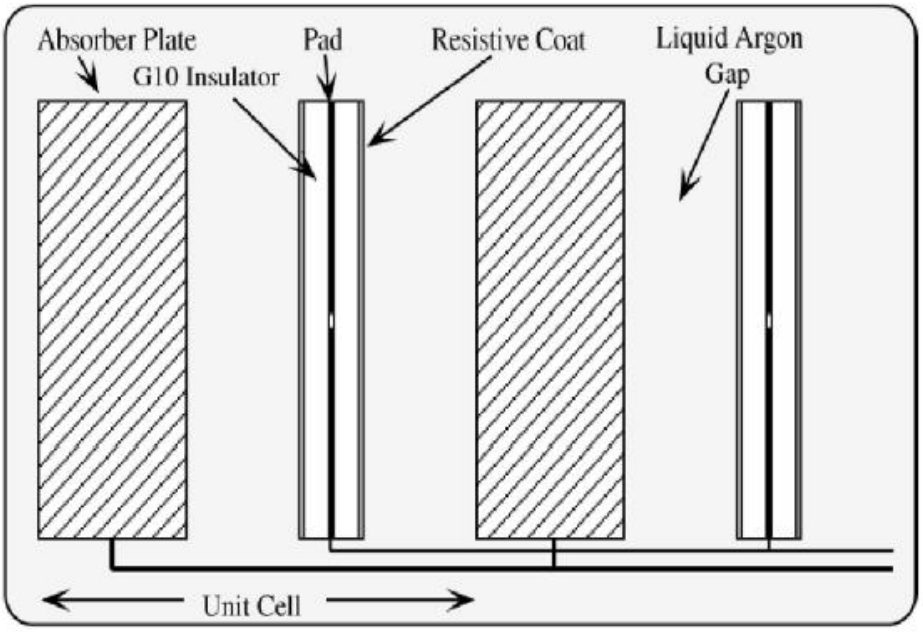


Figure 2.7: Schematic view of two calorimeter cells.

shower.

The EC calorimeter has the same sections as the CC, but is divided into inner, middle, and outer portions. The inner portion contains four EM layers, and several FH and CH layers. The middle portion is made up of FH and CH regions, while the outer portion contains only a CH module.

The calorimeter cells are constructed in a pseudo-projective set of readout towers (Fig. 2.8), with the center of each cell in a given tower lying along the same line to the interaction point, but with the cell boundaries perpendicular to the beam axis. The tower size is generally set at $\Delta\eta = 0.1$, $\Delta\phi = 0.1$.

2.2.5 Muon System

Muons traverse several meters of dense material in the calorimeter without hadronic interactions but leave MIP signals, insufficient for measuring their energy. To detect muons DØ has two types of muon detectors [46] outside the calorimeter: scintillators and drift chambers. Both detectors have three layers which are labeled A, B, and C, from the

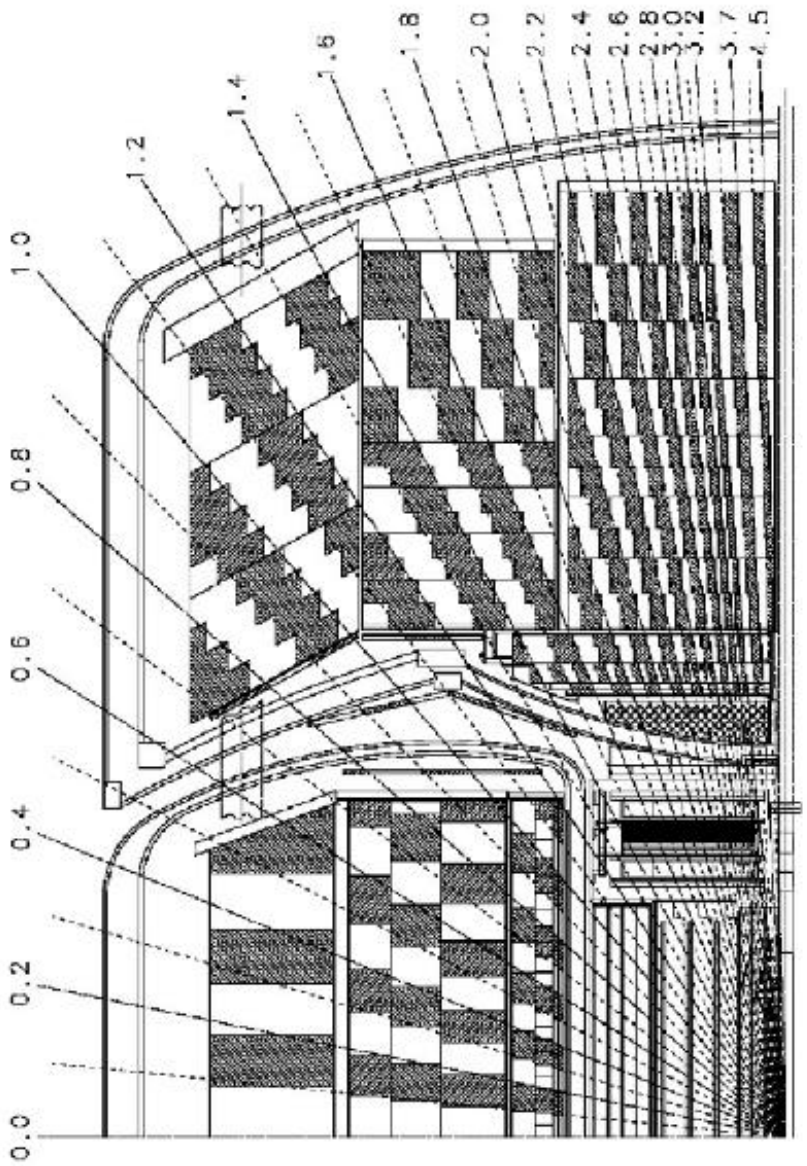


Figure 2.8: Schematic view showing the calorimeter segmentation pattern. The shading pattern indicates cells for signal readout. The radial lines show the detector pseudo-rapidity intervals.

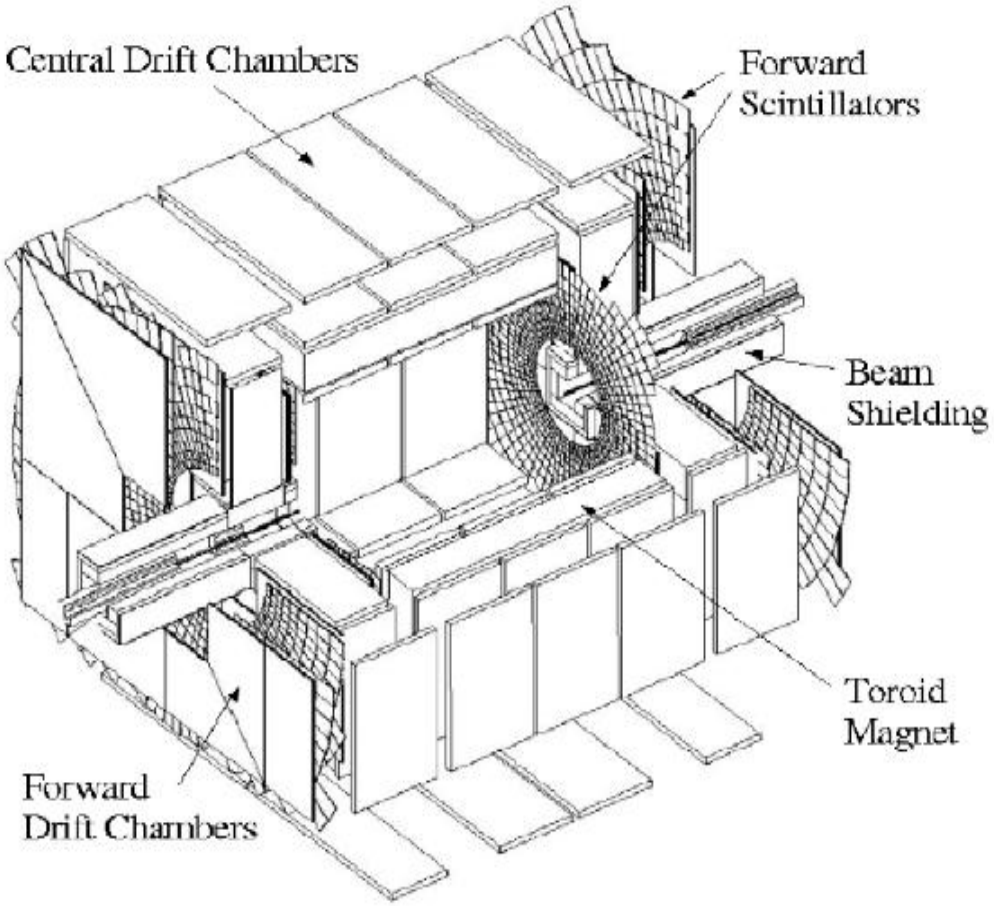


Figure 2.9: The DØ muon system.

innermost to the outermost. The scintillators provide precision timing measurements by triggering on events involving a muon. The drift chambers measure the position of the muon and its momentum. Between the first and second layers of the muon detectors, a 109 cm, 1.8 T toroidal magnet is situated to aid measurement of muon momenta.

The coverage of the system extends to $|\eta| < 2$, split into two regions at $|\eta| < 1$ called the central region and at $1 < |\eta| < 2$ called the forward region. (Fig. 2.9) illustrates the layer structure and separation into the central and forward regions.

Scintillation Counters There are two layers of scintillation counters in the central region, called the Aϕ layer within the toroid, and a cosmic layer outside the toroid. Each scintillation

counter has four walls surrounding the detector and produces a fast readout for triggering and associating muons with the correct bunch crossing. The cosmic layer is also used to reject cosmic ray muons.

The forward region of the muon system [47] includes three layers of scintillation counters. The A layer is within the toroid, while the B and C layers lie outside the toroid. Each scintillation layer is covered with a layer of scintillator pixels which cover a surface of 4.5° in ϕ and 0.1 in η.

Drift Tubes The central muon system includes proportional drift tubes (PDTs) arranged in three layers. The A layer is made up of four decks (sublayers) and located between the calorimeter and the toroid. The B, C layers contain three decks and are placed outside the toroid. The PDTs use a gas mixture of 84% argon, 8% methane, and 8% CF_4, and provide a maximum drift time of 500 ns. Because of the support structures of the detector, drift tube coverage in the central bottom region is incomplete. However, about 90% of the central region is covered by two layers of PDTs and nearly 55% with all three layers. Muon positions can be measured within PDTs with an uncertainty of 1 mm.

The forward region makes use of three layers of mini drift tubes (MDTs) consisting of 1 cm^2 cells. The MDTs, which exhibit more radiation hardness and a faster response time, are new in RunII while the central system reuses PDTs from RunI. Four (for the A layer) or three (for the B and C layers) decks of MDTs are mounted along the field lines of the toroid. As is the case with the central system, the A layer sits before the toroid and the B,C layers are after the toroid. The MDTs use different gas from the PDTs, a 90%-10% mixture of CF_4-CH_4, and have a maximum drift time of 60 ns. The hit position resolution from the MDTs is about 0.7 mm.

Each forward system includes shielding around the beam pipe. The shielding is made of a 51 cm thick slab of iron to absorb hadronic and electromagnetic radiation, followed by 15 cm of polyethylene to absorb neutrons, and 5 cm of lead to block gamma rays.

2.2.6 Luminosity Monitor

The luminosity monitor (LM) is designed to provide precise Tevatron luminosity at the DØ interaction region. Accurate measurement of luminosity is very important because the precision of a cross section measurement depends on it. The secondary purpose of the LM

is to provide diagnostic information of accelerator performance.

The LM is located between the EC calorimeter and the FPS detectors at z = ±140 cm and consists of 24 wedges of scintillator read out by photomultipliers. The coverage extends from 2.7 to 4.4 in pseudorapidity. Particles from collisions create hits at each set of scintillators approximately in coincidence. These coincidences provide for a counter that fires on any beam crossing with a $p\overline{p}$ collision, and beam halo products passing through the detector will appear distinctly separated.

The fundamental unit of time for measuring luminosity is called a luminosity block and each block is indexed by a luminosity block number (LBN) which increases monotonically.

2.2.7 DØ Trigger

Proton antiproton collisions take place every 396 ns at DØ, a rate of over 2.5 MHz and yields a single hard scatter on average. Even though all the DØ detectors have fast readouts, they are still not fast enough to read all events occuring inside the detector. Besides the probabilities of interesting events in the detector are very low, and the DØ data acquisition (DAQ) system is only capable of writing data at approximately 50 Hz to the data storage. All these factors require an efficient trigger system. The DØ trigger system consists of three levels with each imposing selection criteria (cuts) on physics objects. The Level 1 (L1) trigger system is a set of hardware based triggers from individual subdetectros. Events that fulfill L1 requirements are sent to the Level 2 (L2). The L2 uses both software and hardware. Here, basic physics objects are constructed and the first global decision is made. In Level 3, event reconstruction is done based on software algorithms only.

Level 1 trigger The central tracking, calorimeter, and muon systems have their own L1 triggers that search for objects consistent with detector signatures of elementary particles. The L1 central track trigger (CTT) using the CFT and central preshower readout while the L1 muon trigger looks for a charged particle track with momentum exceeding a given threshold. The L1 calorimeter trigger looks for energy deposition exceeding a given threshold. The combined L1 triggers provide a trigger event rate of 1.6 kHz.

Level 2 trigger The L2 system has two distinct stages. The first stage called the preprocessor stage identifies objects from individual detector subsystems, as does the L1

system. In the second stage (global stage), event-wide triggers based on the combined data from all subsystems are generated. Here, the correlation between objects such as leptons and tracks is examined first time. The L2 reduces the trigger rate to 1kHz.

Level 3 *trigger* The L3 trigger is implemented entirely in software that provides limited event reconstruction and makes a trigger decision using all the event information. Approximately 100 computer (farm) nodes handle the job. Each node collects a fully digitized event that passes the L2 trigger and provides a final trigger decision. Algorithms are used to create and cut on the individual objects, such as a minimum momentum for muons or non-physics objects, such as vector sums of transverse energy. Finally, events meeting the L3 trigger requirements are written to tape at a rate of 50 Hz. Each event is approximately 50 Mbytes in size.

CHAPTER 3

Event Reconstruction

The DØ detector is designed to distinguish and define fundamental objects such as tracks, electrons, muons, and jets. However, the raw data from the detector are just digitized pulse height and time information. These data must be processed by event reconstruction algorithms to produce physics objects and their properties, representing the particles that originated from a $p\overline{p}$ collision. The collection of reconstruction algorithms is referred to as *d0reco* and this process is called object identification.

Event reconstruction is also required for Monte Carlo (MC) simulated events. This analysis uses signal and background generated from MC techniques as input of Bayesian neural networks. Therefore, agreement between real (often referred to as "data") and MC events is very important.

This chapter discusses how the detector subsystems are used to reconstruct physics objects, and gives a description of corrections which make MC events agreement with real data, on the average.

3.1 Tracks

Charged particles traversing the SMT and CFT detectors leave traces in the form of small energy deposits in many layers, called hits. The DØ tracking algorithm reconstructs a particle track from the collection of hits in any given event. Since there are tracks from secondary collisions, random electric noise, and an event producing many charged particles in a small range of η, it is very difficult to recognize which hit is formed from which track.

There are several approaches to reconstruct tracks from hits: histogramming track finding (HTF), alternative algorithm (AA), and a combination of the HTF and AA called global track reconstruction (GTR).

Histogramming Track Finding (HTF)

The HTF method [48] finds SMT hits or CFT hits, maps the hits in x-y coordinates into a new parameter space defined by coordinates ϕ-ρ (where ϕ is the azimuthal angle, and ρ the curvature of a track). In the ϕ-ρ space, a track candidate will produce a peak and random hits will be distributed uniformly.

A 2D Kalman filter, which is similar to a least square fit, fires to identify real tracks. The second transformation adds z coordinate information and is performed from r-z space to z_o-C space (where z_o is the intersection of the track along the beam axis and C is the track inclination defined as dz/dr). A peak is found again in the z_o-C space, and the 3D Kalman filter is used to fit the remaining hits. Finally, the newly formed tracks are extrapolated.

Alternative Algorithm (AA)

The AA algorithm begins by generating a pool of track candidates using the hits in the SMT. The algorithm selects all sets of three hits which lie along a path originating from the beam spot. It then extrapolates the path of the track outward to either the next layer of the SMT or to the CFT to calculate the point where the track should have crossed the next layer. The algorithm checks whether there is a hit near that location, and then extrapolates to the next layer, and repeats the procedure. At each layer a χ^2 of hits with respect to the track is calculated and the hit becomes associated with the track if its χ^2 is less than a certain value. If there is no hit in the layer, the algorithm continues and records a "miss" for this track. At the end of this procedure, a list of tracks is produced along with hits, misses, and χ^2-s.

Next, a list of vertices is constructed from tracks propagated back to the beam axis. These vertices are used to look at the track candidates that have only CFT hits. The same extrapolation procedure is used but starting in the CFT with the constraint that tracks originated at one of the vertices. CFT-only tracks have been used in several analyses and provide higher efficiency, although at some expense of resolution.

Global Track Reconstruction (GTR)

The GTR algorithm uses the reconstructed tracks from both the HTF and AA algorithms as inputs. A standard Kalman fit is applied to these tracks, and the final set of tracks in the event is defined.

3.2 Primary Vertices

The primary vertex (PV) [49] is defined as the initial $p\overline{p}$ collision point of the hard scattering in three dimension. The precise determination of the PV is critical for reconstruction of physics objects, and for discrimination against low energy inelastic $p\overline{p}$ collisions.

The location of the PV is reconstructed by means of an adaptive primary vertex algorithm. This algorithm consists of three main steps.

Track Selection

Tracks with $p_T > 0.5$ GeV and at least two SMT hits are assigned to a vertex.

Vertex Fitting

In the first pass, χ^2 for each track hypothesis is computed. Tracks providing the largest contributions to the χ^2 are removed one by one until the overall χ^2 is less than 10.

In the second pass, the track *dca* (transverse distance of closest approach) significance is calculated with respect to the position of these first-pass vertices. Only tracks with less than 5σ ($dca/\sigma(dca) < 5$) are fitted to the final PV.

Vertex Selection

The final vertex is calculated from the remaining tracks. In the case where more than one vertex is found, the p_T distributions of the tracks associated with each vertex are used to define a probability that each track originated at the particular vertex. The vertex with the largest weighted product of track probabilities, which has the lowest minimum bias probability, is identified as the PV.

3.3 Electrons

An electron deposits energy in the calorimeter and also makes a track in the central tracking system [50]. Information from the two subdetectors is used to reconstruct an electron, which proceeds in two stages. The first stage involves a cluster formed in the calorimeter. At the second stage, confirmation is sought from the central tracker.

A simple cone algorithm forms electromagnetic clusters of towers in the electromagnetic section of the calorimeter. These towers are included within a cone of radius 0.4 around

the highest energy tower. Since an electron is expected to deposit most of its energy in the first layers of the calorimeter, genuine EM showers will have a large EM fraction $EMF = E_{EM}/E_{tot}$ where E_{EM} is the cluster energy in the EM section of the calorimeter and E_{tot} is its total energy within the cone.

To test the consistency of the cluster's shower with the hypothesis that it is an electron, each shower is attributed a χ^2_{cal} based on seven variables comparing the values of the energy deposited in each layer of the EM calorimeter and the total energy of the shower with expected shower profiles from the MC simulation. Additionally, the cluster is required to be isolated. The isolation

$$f_{iso} = \frac{E_{tot}(\Delta R < 0.4) - E_{EM}(\Delta R < 0.2)}{E_{EM}(\Delta R < 0.2)}, \tag{3.1}$$

where $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, must be small. Finally, an electron likelihood [51] is defined based on seven variables including both calorimeter and tracking information. In this analysis, the following selection criteria are used:

- $EMF > 0.9$
- $\chi^2_{cal} < 50$
- $f_{iso} < 0.15$
- Likelihood > 0.85

3.4 Muons

Muons do not produce showers in the calorimeter. Therefore, they are reconstructed using information from the muon system and the central tracking detector [52]. The muon system can identify muons with high purity solely from tracks. To improve the momentum resolution, the track in the muon system is required to be matched with a track found by the tracking system. The starting point for muon reconstruction is the formation of straight line track segments from hits in each layer of the muon system. After that, conforming scintillator hits in the same layer are matched. If A-layer segments (inside the toroid) and BC-layer segments (outside the toroid) are consistent, it is possible to make momentum

measurement from the curvature induced by the toroid magnet. Muon candidates formed this way are called local muons.

To describe the type of segments found in an event, a variable nseg is defined. A muon with an A-layer has |nseg| = 1, a muon with a B or C-layer has |nseg| = 2, and a muon with both segments has |nseg| = 3. The nseg value is positive if the muon segment is matched to a central track.

Muons are also categorized depending on the location and types of hits in the muon system. These classifications are referred to as "tight", "medium", and "loose". Tight muons require drift tube and scintillator hits both inside and outside the toroid, and result from a local muon track match. Medium muons require drift tube and scintillator hits out side the toroid. In the bottom region, where muon detector coverage is reduced, this requirement is relaxed. Loose muons require only one reconstructed segment. More detailed description of the muon classification is given in Ref. [53].
In this analysis, the requirements for muon reconstruction are as follows:

- At least two A-layer wire hits
- At least one A-layer scintillator hit
- At least two wire hits in the B and C-layers combined
- At least one scintillator hit in the B and C-layers combined
- A matched track to the central tracking systems, SMT and CFT

To remove muons produced by cosmic rays, we require

- A, B, and C-layer scintillation times < 10 ns from the bunch crossing time
- $dca < 0.2$ cm if there are no SMT hits
- $dca < 0.02$ cm if there is at least one SMT hit
- Δz(central track, PV) < 1 cm

Finally, muons must be well isolated. This requirement favors a muon that has come from W boson decay, and not from heavy flavor decay. To remove muons from heavy flavor decays, we require

- ΔR(μ, jet) > 0.5
- Track halo isolation $= \sum_{\Delta R<0.5} p_T^{trk}/p_T^{\mu} < 0.2$
- Calorimeter halo isolation $= \sum_{0.1<\Delta R<0.4} E_T^{cell}/p_T^{\mu} < 0.2$

3.5 Jets

Quarks and gluons cannot exist alone as free particles. They hadronize into many colorless particles due to color confinement. These particles are detected as showers within a narrow cone in the calorimeter, called jets [54].

Since jets are not fundamental objects, they must be modeled. This is done by DØ in RunII using the T42 algorithm and the improved legacy cone algorithm (ILCA) [53]. The T42 algorithm removes noisy cells before reconstructing the calorimeter objects. The algorithm accepts an isolated cell as a "signal-like" cell if the energy deposited in the cell is above a threshold of 4σ, or above 2.5σ and in addition the energy of a neighboring cell is above 4σ, where σ is the mean width of noise in the cell.

Once noisy cells are removed from the towers seeds are founded, that is represented by calorimeter towers above a minimum seed threshold. The ILCA selects calorimeter towers with transverse energy > 0.5 GeV as seeds, and collects all calorimeter towers within a cone of radius ΔR $= 0.5$ around the seed. If the cone has transverse energy greater than 1GeV, it is defined as a jet candidate. The central axis of the jet candidate is an E_T weighted midpoint of the towers. Next, a new cone is drawn around the new direction and the towers within the new cone are used to calculate the new central axis of the jet candidate, which is compare to the central axis of the old jet candidate. This process continues until the jet axis does not change appreciately or the maximum number of iterations (50) is reached. In the final step, the jet finding algorithm decides whether to merge or split jets that share energy. If the shared energy is less than 50% of the individual jet energies, jets are split into two distinct ones, otherwise they are merged.

After the jets have been reconstructed, criteria (quality cuts) are applied to distinguish electromagnetic particles from real jets. The specific selection criteria are summarized below.

- To remove electromagnetic particles a cut on the fraction of the total calorimeter energy

in the jet cone contained in the EM region is applied at $0.05 < f_{EM} < 0.95$, where $f_{EM} = E_{EM}/E_{total}$

- Since the noise level is higher in the coarse hadronic calorimeter, a cut on the fraction of the jet energy deposited in the coarse hadronic calorimeter is applied at $f_{CH} < 0.4$, where
  $f_{CH} = E_{CH}/E_{total}$. This cut is designed to remove fake jets which are clustered around noise.

- A jet is required to be isolated from all electromagnetic clusters, ΔR > 0.5.

- To remove jets clustered partially from noisy cells (hot cells), a cut on the ratio of the energy in the most energetic cell of a jet to that of the second most energetic cell (f_{hot}) is applied at $f_{hot} < 10$.

- n90 is defined as the number of calorimeter towers in a jet that contain 90% of the total energy of the jet. To remove jets clustered from a single hot hower, n90 is required to be greater than 1.

Jet Energy Scale

The jet energy measured in the calorimeter is not equal to the energy of the parton from which the jet is presumed to have arisen. This is due to non-linearities, noise, dead material, and showering effects in the calorimeter. Therefore, the energy of the calorimeter jet must be corrected to estimate the parton energy.

The DØ experiment uses the jet energy scale (JES) algorithm to correct the detector response: JES attempts to correct the reconstructed jet energy, E^{reco}, back to the particle level energy that is the jet energy before the interaction with the calorimeter, E^{corr} [55]. The correction can be written as

$$E^{corr} = \frac{E^{reco} - O}{R \times S}, \tag{3.2}$$

- **Energy Offset O:** energy in the clustered cells from electronic noise, pile-up, the underlying event, multiple interactions, and noise from radioactive decay of the uranium in the calorimeter. The offset correction is measured by averaging over

minimum bias data samples and averaging over ϕ, and is given as a function of detector η (that is, y measured from the coordinant origin) and instantaneous luminosity.

- **Calorimeter Response R:** R is a measure of the calorimeter response to a jet. It is less than one because of energy loss in the detector before the calorimeter and the poorer response of calorimeters to hadrons than electrons. In addition, the measured jet energy can be distorted due to the non-linear response of the calorimeter to the particle energies, a different response of the calorimeter to different particles, and un-instrumented regions of the detector or dead material. This response (R) is determined using the transverse energy balance in back-to-back photon+jet events. The transverse energy of the photon is measured very precisely and provides the target transverse energy for the jet. This correction is derived as a function of the jet energy and detector η.

- **Showering Corrections S:** This is a measure of the energy that radiates outside of the cone during the shower development in the calorimeter. Furthermore, the solenoid field can change a particle's trajectory. The correction is determined from the jet profiles in the photon+jet sample by comparing the transverse energy balance for varying cone radii.

3.6 b Quark Jets

Identifying jets from the hadronization of b quarks is very important because many interesting particles such as top quarks decay into a b quark (almost 100%). Hadrons containing b quarks have much longer life time than light hadrons. They can travel a few milimeters in the detector before decaying. The identification of jets from b quarks is refered to as b tagging. DØ develped a few different algorithms to distinguish b jets from light jets [56]. One, the neural netwok b tagging algorithm, combines seven variables from the existing algorithms and provides the best b tagging performance. Variables used in the network are listed below.

- Decay length significance of the Secondary Vertex

- Weighted combination of the tracks' impact parameter significance

- Probability that the jet originates from the primary vertex
- χ^2 per degree of freedom of the secondary vertex
- Number of tracks used to reconstruct the secondary vertex
- Mass of the secondary vertex
- Number of secondary vertex found in the jet

3.7 Missing Transverse Energy

Since a neutrino is the least interacting particle, it does not leave any trace in the detector. The presence of a neutrino in the final state can be detected only from the imbalance of the energy in the transverse plane. The missing transverse energy, $\not\!\!E_T$, is defined as the vector sum of the energies deposited in the fine hadronic and electromagnetic calorimeter cells minus the lepton p_T vector such that there is no net transverse momentum in the event.

CHAPTER 4

Data Samples

The data sample was collected between August 2002 and December 2005, with run numbers 151,817–213,063 inclusive. We have nearly 1 fb^{-1} of good quality data in each of the electron and muon channels, as shown in Tables 4.1 and 4.2.

The data we are using have been reconstructed with version p17.09.03 of the DØ production code. They have been obtained from the DØ Common Samples Group's data skims.

Table 4.1: Integrated luminosities by trigger version for the electron channel.

		Electron Channel Integrated Luminosity [pb^{-1}]		
Trigger Version	Trigger Name	Delivered	Recorded	Good Quality
v8.0 – v9.0	EM15_2JT15	6	5	5
v9.0 – v10.0	EM15_2JT15	48	42	25
v10.0 – v11.0	EM15_2JT15	20	18	10
v11.0 – v12.0	EM15_2JT15	79	72	63
v12.0 – v13.0	E1_SHT15_2J20	273	251	227
v13.0 – v13.3	E1_SHT15_2J_J25	80	73	55
v13.3 – v14.0	E1_SHT15_2J_J30	354	325	294
v14.0 – v15.0	E1_SHT15_2J_J25	290	271	234
Total Integrated Luminosity		1,150	1,056	**913**

The electron channel trigger requires at least one electron and at least two jets. The muon channel trigger requires one muon and at least one jet. A description of the electron and muon triggers used is given below.

Table 4.2: Integrated luminosities by trigger version for the muon channel.

Muon Channel Integrated Luminosity [pb^{-1}]				
Trigger Version	Trigger Name	Delivered	Recorded	Good Quality
V8.0 - V9.0	MU_JT20_L2M0	7	6	6
V9.0 - V10.0	MU_JT20_L2M0	48	42	25
V10.0 - V11.0	MU_JT20_L2M0	21	19	11
V11.0 - V12.0	MU_JT20_L2M0	79	74	65
V12.0 - V13.0	MU_JT25_L2M0	277	255	231
V13.0 - V13.2	MUJ2_JT25	56	39	31
V13.2 - V13.3	MUJ2_JT25_LM3	26	22	16
V13.3 - V14.0	MUJ2_JT30_LM3	382	277	252
V14.0 - V14.2	MUJ1_JT25_LM3	0	0	0
V14.2 - V14.3	MUJ1_JT25_ILM3	25	23	21
V14.3 - V15.0	MUJ1_JT35_LM3	265	248	214
Total Integrated Luminosity		1,187	1,006	**871**

Electron Trigger

- EM15_2JT15
  - Level1: One EM calorimeter tower with $E_T > 10$ GeV and two jet calorimeter towers with $E_T > 5$ GeV.
  - Level2: One EM object with $E_T > 10$ GeV and electromagnetic fraction > 0.85. Also two jet objects with $E_T > 10$ GeV.
  - Level3: One EM object with $E_T > 15$ GeV and a shower shape consistent an EM object. Also, two jet objects with $E_T > 15$ GeV.
- E1_SHT15_2J20
  - Level1: One EM calorimeter tower with $E_T > 11$ GeV.
  - Level2: No requirement.
  - Level3: One EM object with $E_T > 15$ GeV and a shower shape consistent an EM object. Also, two jet objects with $E_T > 20$ GeV.
- E1_SHT15_2J_J25

- Level1: One EM calorimeter tower with $E_T > 11$ GeV.
- Level1: One EM object with $E_T > 15$ GeV.
- Level2: No requirement.
- Level3: One EM object with $E_T > 15$ GeV and a shower shape consistent an EM object. Also, two jet objects with $E_T > 20$ GeV. One of the jets is also required have $E_T > 25$ GeV.

- E1_SHT15_2J_J30
  - Level1: One EM calorimeter tower with $E_T > 11$ GeV.
  - Level1: One EM object with $E_T > 15$ GeV.
  - Level2: No requirement.
  - Level3: One EM object with $E_T > 15$ GeV and a shower shape consistent an EM object. Also, two jet objects with $E_T > 20$ GeV. One of the jets is also required have $E_T > 30$ GeV.

Muon Trigger

- MU_JT20_L2M0
  - Level1: One muon with scintillator and wire hit and one calorimeter tower with $E_T > 5$ GeV.
  - Level2: One muon object.
  - Level3: One jet object with $E_T > 20$ GeV.

- MU_JT25_L2M0
  - Level1: One muon with scintillator and wire hit and one calorimeter tower with $E_T > 5$ GeV.
  - Level2: One muon object.
  - Level3: One jet object with $E_T > 25$ GeV.

- MUJ2_JT25

  - Level1: One muon with scintillator and wire hit and one calorimeter tower with $E_T > 5$ GeV.
  - Level2: One muon object and a jet object with $E_T > 8$ GeV.
  - Level3: One jet object with $E_T > 25$ GeV.

- MUJ2_JT25_LM3
  - Level1: One muon with scintillator and wire hit and one calorimeter tower with $E_T > 5$ GeV.
  - Level2: One muon object and a jet object with $E_T > 8$ GeV.
  - Level3: One jet object with $E_T > 25$ GeV and a muon object with pT > 3 GeV.

- MUJ2_JT30_LM3
  - Level1: One muon with scintillator and wire hit and one calorimeter tower with $E_T > 5$ GeV.
  - Level2: One muon object and a jet object with $E_T > 8$ GeV.
  - Level3: One jet object with $E_T > 30$ GeV and a muon object with pT > 3 GeV.

- MUJ1_JT25 ILM3
  - Level1: One muon with scintillator and wire hit and one calorimeter tower with $E_T > 5$ GeV.
  - Level2: One muon object and a jet object with $E_T > 8$ GeV.
  - Level3: One jet object with $E_T > 25$ GeV and an isolated muon object with pT > 3 GeV.

- MUJ1_JT35_LM3
  - Level1: One muon with scintillator and wire hit and one calorimeter tower with $E_T > 5$ GeV.
  - Level2: One muon object and a jet object with $E_T > 8$ GeV.
  - Level3: One jet object with $E_T > 35$ GeV and a muon object with pT > 3 GeV.

CHAPTER 5

Signal and Background Estimation

Our search for single top quark production relies on simulated single top quark and background events.

5.1 Single Top Event Generation

Single top (signal) events are generated using the CompHEP-SingleTop Monte Carlo event generator [57]. SingleTop produces events whose kinematic distributions match those from NLO calculations. The top quark mass is set to 175 GeV, the parton distribution function set (pdf set) is CTEQ6L1, and the scales are m_t^2 for the s-channel and $(m_t/2)^2$ for the t-channel. The top quarks and the W bosons from the top quark decays were decayed in CompHEP-SingleTop to ensure the spins are properly transfered. PYTHIA [58] was used to add the underlying event and initial- and final-state radiation. TAUOLA [59] was used to decay tau leptons, and EVTGEN [60] to decay b hadrons.

5.2 Modeling Backgrounds

The backgrounds to the single top signal can be separated into two categories, physics backgrounds and instrumental backgrounds. Physics backgrounds share the same event signature as signal events, with an isolated high p_T lepton, two or more jets, and missing transverse energy, $\not\!\!E_T$. The largest physics background is the production of a W boson together with two or more jets. This background is referred to as "W+jets".

Another physics background is top pair production, $t\bar{t}$. This background includes two W bosons from the decay of the two top quarks. When both W bosons decay leptonically, which gives rise to "dilepton" events, two quarks, two leptons, and two neutrinos are present

in the final state. The other case where one of the W bosons decays hadronically and the other decays leptonically produces "lepton+jets" events.

The instrumental background is multijet events produced by the strong interactions. Its event signature differs from that of the single top signal in that it has no real missing transverse energy from neutrinos. However, event misreconstruction may give rise to enough fake $\not\!\!E_T$ to allow some of these events to pass the selection criteria. In case of electron events, one of the reconstructed jets has a large electromagnetic fraction causing it to be mis-identified as an electron. Muon events have a gluon splitting into a $b\bar{b}$ pair in which one of the B mesons decays semi-leptonically and produces a muon.

5.3 Background Generation

5.3.1 Top Pair Production

The $t\bar{t}$ samples are generated using ALPGEN [61] and PYTHIA for subsequent generation of final state radiation, fragmentation and decay into stable particles. The top quark mass is set to 175 GeV, the renormalization and factorization scale Q^2 is $m_t^2 + P_T^2(jets)$ and the pdf set used is CTEQ6L1.

Table 5.1: Event numbers, cross section, and weights for $t\bar{t}$ MC events.

Sample	Events	Cross Section [pb]	Weight
$t\bar{t}\rightarrow l + jets$+0lp	481572	71.15	0.039
$t\bar{t}\rightarrow l + jets$+1lp	336400	29.85	0.036
$t\bar{t}\rightarrow l + jets$+2lp	332347	10.25	0.016
$t\bar{t}\rightarrow ll$+0lp	738761	19.18	0.014
$t\bar{t}\rightarrow ll$+1lp	161300	7.939	0.011
$t\bar{t}\rightarrow ll$+2lp	171411	2.636	0.005

Table 5.1 shows event numbers, cross section, and weights for all $t\bar{t}$ samples.

5.3.2 W+jets

In the W+jets sample generation, ALPGEN is used for the hard scatter calculation and PYTHIA for subsequent generation of final state radiation, fragmentation and decay into stable particles. The set of parton distribution functions used are CTEQ6L1. The parton-level selection on the lepton is $|\eta| < 10$, while the selections for jets are $p_T > 8$GeV and $|y| < 3.5$. The minimum distance between two jets is $\Delta R(j_1, j_2) = 0.4$, while no cut is applied on the minimum distance between a jet and the lepton. The renormalization and factorization scale is $m_W^2 + P_T^2(W)$. EVTGEN is used to provide the various branching fractions and lifetimes for B-hadrons.

To get a reliable estimate of the flavor composition, separate samples are generated for the various combinations of quark flavors. The samples are classified as follows:

- W + light jets, events with light flavor partons ($udsg$) as well as W+c+light flavor events where c quark is considered massless
- W + $c\bar{c}$ + light jets, events with two c quarks which are massive
- W + $b\bar{b}$ + light jets, events with two b quarks

When generating events in each jet multiplicity bin, a matching of partons and jets is necessary in order to eliminate the double counting of events. This matching procedure also reduces the sensitivity of the parton-level cross sections, predicted by ALPGEN, to the parton generation cuts. The matching procedure used in this analysis is called MLM matching, proposed and named after M. L. Mangano [62].

MLM Jet-Parton Matching

Events that do not satisfy the following conditions, are rejected from the MC samples:

- The number of jets before hadronization is required to be equal to the number of partons
- Jets are required to be matched if they have $p_T > 15$ GeV and there is a parton with $\Delta R < 0.7$ from the jet.

Table 5.2 shows event numbers, cross section, and weights for all W+jets samples.

Table 5.2: Event numbers, cross section, and weights for W+jets MC events.

Sample	Events	Cross Section [pb]	Weight
W+0lp	7844750	4574	2.15
W+1lp	1053000	1273	0.68
W+2lp	1250500	298.5	0.34
W+3lp	621000	70.56	16.4
W+4lp	582250	15.83	0.07
W+5lp	41750	11.29	0.13
$W+c\bar{c}$+0lp	481572	71.15	0.039
$W+c\bar{c}$+1lp	336400	29.85	0.036
$W+c\bar{c}$+2lp	332347	10.25	0.016
$W+c\bar{c}$+3lp	372248	18.39	0.020
$W+b\bar{b}$+0lp	738761	19.18	0.014
$W+b\bar{b}$+1lp	161300	7.939	0.011
$W+b\bar{b}$+2lp	171411	2.636	0.005
$W+b\bar{b}$+3lp	163674	1.742	0.003

5.3.3 Multijet Events

Multijet production is the only instrumental background in this analysis. The multijet background is modeled using real events which pass all selection cuts except the isolation cut for muons or likelihood cut for electrons.

5.4 Trigger Simulation

Monte Carlo modeling of the DØ trigger system can be done either by using the DØ TrigSim program, or by folding into Monte Carlo simulated events the per-electron, per-muon and per-jet probability of satisfying individual trigger conditions at Level 1 (L1), Level 2 (L2) and Level 3 (L3) [53]. The probability of a single object (electron, muon, jet)

to satisfy a particular trigger requirement is measured in data. Single object efficiencies are in general parameterized as a function of the kinematic variables (p_T, η, and ϕ) of the online reconstructed objects.

Although correlations and overlap between triggers are automatically taken into account using the first method, currently, the Monte Carlo modeling of trigger objects and trigger quantities is not precise enough to be reliable. Therefore the second method based on trigger efficiencies derived from data is used in this analysis.

The approach used to combine single object trigger efficiencies to calculate the probability of an event to satisfy a specific trigger is described in Ref. [63] and briefly summarized below. The total event probability is calculated as the product of the probabilities for the event to satisfy the trigger conditions at each triggering level,

$$P(L1, L2, L3) = P(L1)P(L2|L1)P(L3|L1, L2), \tag{5.1}$$

where $P(L2|L1)$ and $P(L3|L1, L2)$ represent the conditional probability for an event to satisfy a set of criteria given it has already passed the requirements imposed at the previous triggering level(s). Conditional probability is defined by

$$P(a|b) = \frac{P(a, b)}{P(b)}. \tag{5.2}$$

The total probability of an event to satisfy a set of trigger requirements is obtained assuming that the probability for a single object to satisfy a specific trigger condition is independent of the presence of other objects in the event. Under this assumption, the contributions from different types of objects to the total event probability can be written as

$$P_{(object1\ and\ object2)} = P(object1)P(object2). \tag{5.3}$$

Furthermore, under this assumption, the probability (P) for at least one object to satisfy a particular trigger condition, out of a total of N objects present in an event, is given by

$$P = 1 - \prod_{i=1}^{N}(1 - P_i), \tag{5.4}$$

where P_i represents the single object probability (the probability for an electron to fire a jet trigger and vice versa is also considered in the corresponding product).

The total trigger efficiency is calculated as the luminosity weighted average of the event probability associated with the trigger requirements contained in each trigger list.

The probability of a single object to satisfy a particular trigger requirement is measured using the following general procedure. The first step consists of identifying a sample of events, unbiased with respect to the trigger requirement under study. Offline reconstructed objects are then identified in the events. The efficiency is obtained by calculating the fraction of these offline reconstructed objects that satisfy the trigger condition under study.

5.5 Monte Carlo Corrections

Monte Carlo events generated with ALPGEN (or CompHEP) and PYTHIA are processed using the DØGSTAR or GEANT based detector simulator [64]. The resulting output of the Monte Carlo simulations has the same format as the data, and the same reconstruction routines are applied to both. However, the simulated events have to be modified to better describe real events. The efficiencies of the object reconstruction and associated correction factors are discussed in the following sections.

5.5.1 Primary Vertex

The primary vertex efficiency is measured with real and Monte Carlo $Z \to \mu\mu$ events. The correction factor is defined as the ratio of the two efficiencies. In this analysis, no correction factor for the Monte Carlo primary vertices is applied.

5.5.2 Electrons

Electron correction factors are measured using $Z \to ee$ data and Monte Carlo events. One electron is required to pass tight selection cuts and the electron efficiency is measured with the other electron. This is called the tag-and-probe method. The ratios of the data and Monte Carlo efficiencies are used to weight the Monte Carlo events. The correction factor for electrons is considered a product of two independent factors: reconstruction and track match plus likelihood cuts,

$$f_{e-ID} = \frac{\epsilon^{Data}_{Reco}}{\epsilon^{MC}_{Reco}} \times \frac{\epsilon^{Data}_{TrackMatchLikelihood|Reco}}{\epsilon^{MC}_{TrackMatchLikelihood|Reco}}. \tag{5.5}$$

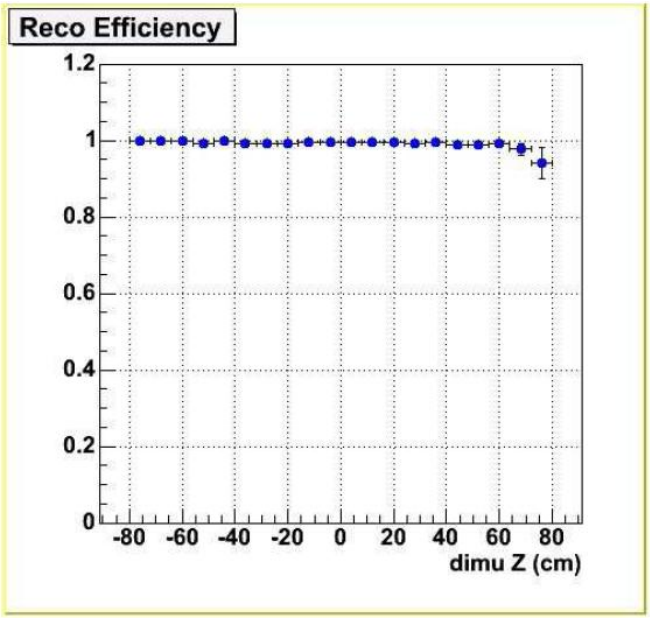


Figure 5.1: Reconstruction efficiency of primary vertex.

5.5.3 Muons

The detector resolution of muon Monte Carlo events is overestimated, while simulated electrons have well-modeled resolution. To account for such overestimation the reconstructed Monte Carlo muons need to be "smeared" to ensure that the resolution matches that of real muons. The momentum smearing can be expressed as

$$\left(\frac{q}{p_T}\right)' \rightarrow \frac{q}{p_T} + \left(A + \frac{B}{p_T}\right) \times G, \tag{5.6}$$

where the parameter G is a random number generated from a Gaussian distribution centered at 0 and a width of 1. The parameters A and B are measured for muons with an SMT track hit in two regions ($\eta < 1.6$ and $\eta > 1.6$) and for muons without an SMT hit. The muon track is defined by the charge and radius of curvature, which is proportional to q/p_T, thus the natural quantity to smear is q/p_T.

After the smearing is applied, the tag-and-probe method described in the previous section is used to measure the muon correction factor which is defined as the product of three independent factors for reconstruction, track matching, and isolation.

$$f_{\mu-ID} = \frac{\epsilon^{Data}_{Reco}}{\epsilon^{MC}_{Reco}} \times \frac{\epsilon^{Data}_{Track|Reco}}{\epsilon^{MC}_{Track|Reco}} \times \frac{\epsilon^{Data}_{Isolation|Track}}{\epsilon^{MC}_{Isolation|Track}}. \quad (5.7)$$

5.5.4 Jets

In jet identification and reconstruction, several data and Monte Carlo corrections are applied to compensate for differences. The main issue is that the jet energy resolution is too good in the MC events and there is a small offset not corrected by the jet energy scale correction. The energies of the Monte Carlo jets are smeared to make their resolution match that of real jets.

5.5.5 Tag Rate Functions

The neural network tagger (see Sec. 3.6) is not used to identify b jets in Monte Carlo events since the tagger and tracking reconstruction in Monte Carlo events are not well modeled. To manage such problems, DØ derived Taggability Functions and Tag-Rate Functions (TRF) for b jet, c jet, and light jet in an event. Taggability Functions estimate the chance that a jet could be tagged. This requires at least two associated tracks with $p_T > 1$ GeV for jets. After the Taggability Functions are applied the TRF is used to calculate the probability that a jet is b tagged.

The TRF for data is measured using two independent data samples. The first data sample is required to have one muon found inside of a jet cone. This requirement is expected to select events including semi-leptonic b decay. The second sample requires at least two jets of which one must have a jet impact parameter probability less than 0.5. The neural network tagger and soft lepton tagger are used to construct a system of eight equations and eight unknowns. The system is then solved yielding tagging efficiencies as a function of E_T and detector η.

The b tagging efficiencies with Monte Carlo events are also simulated. The Taggability Function and TRF are measured as a function of jet p_T and η, and the ratio of data and Monte Carlo efficiency is calculated to obtain the correction factors.

In addition to b tagging efficiency, the neural network tagger measures c jet tagging efficiency using a similar approach. The c jet tagging efficiency is also a function of jet p_T and η.

Light jets are sometimes mistagged by the neural network tagging algorithm. This rate is derived from the negative tag-rate. The negative tag-rate is the probability with which a jet resulting from light flavor partons is mis-identified as a b jet. This is measured in real events with little bias towards heavy-flavor events.

CHAPTER 6

Event Selection

The main purpose of the event selection is to find W-like events containing an isolated lepton, missing transverse energy, and one to four jets with high transverse momentum. The selection cuts applied to the data are described below. Samples after this selection will be used in further analysis steps of sophisticated discrimination techniques.

- Good quality [65] (for data)
- Pass trigger: offline electrons and muons in the data are matched to the object that fired the appropriate trigger for that run period and triggerlist
- Good primary vertex: $|z_{\mathrm{PV}}| < 60$ cm with at least three tracks attached
- Missing transverse energy $15 < \not\!\!E_T < 200$ GeV
- One, two, three or four jets with $p_{\mathrm{T}} > 15$ GeV and $|\eta| < 3.4$
- The leading jet is required to have $p_{\mathrm{T}} > 25$ GeV and $|\eta| < 2.5$
- The second leading jet is required to have $p_{\mathrm{T}} > 20$ GeV
- There must be fewer than three noise jets
- Jet triangle cut $|\Delta\phi(\text{leading jet}, \not\!\!E_T)|$ vs. $\not\!\!E_T$: from 1.5 to π radians when $\not\!\!E_T = 0$ GeV, and $\not\!\!E_T$ from 0 to 35 GeV when $|\Delta\phi| = \pi$ rad

The electron channel selection:

- Only one tight electron with $E_T > 15$ GeV and $|\eta^{\mathrm{det}}[1]| < 1.1$

[1] $\eta^{\mathrm{det}} = -\ln\left[\tan\frac{\theta^{\mathrm{det}}}{2}\right]$ where θ^{det} is the polar angle as measured from the center of the detector.

- No second loose electron with $E_T > 15$ GeV
- No tight muon with $p_{\mathrm{T}} > 18$ GeV and $|\eta^{\mathrm{det}}| < 2.0$
- Electron coming from the primary vertex: $|\Delta z(e, \mathrm{PV})| < 1$ cm
- Electron triangle cuts $|\Delta\phi(\mathrm{e}, \not\!\!E_T)|$ vs. $\not\!\!E_T$:
  - from 2 to 0 rad when $\not\!\!E_T = 0$ GeV, and $\not\!\!E_T$ from 0 to 40 GeV when $|\Delta\phi| = 0$ rad
  - from 1.5 to 0 rad when $\not\!\!E_T = 0$ GeV, and $\not\!\!E_T$ from 0 to 50 GeV when $|\Delta\phi| = 0$ rad
  - from 2 to π rad when $\not\!\!E_T = 0$ GeV, and $\not\!\!E_T$ from 0 to 24 GeV when $|\Delta\phi| = \pi$ rad

The muon channel selection:

- Only one tight muon with $p_{\mathrm{T}} > 18$ GeV and $|\eta^{\mathrm{det}}| < 2.0$
- No tight electron with $p_{\mathrm{T}} > 15$ GeV and within $|\eta^{\mathrm{det}}| < 2.5$
- Muon coming from the primary vertex: $|\Delta z(\mu, \mathrm{PV})| < 1$ cm
- Muon triangle cuts $|\Delta\phi(\mu, \not\!\!E_T)|$ vs. $\not\!\!E_T$:
  - from 1.1 to 0 rad when $\not\!\!E_T = 0$ GeV, and $\not\!\!E_T$ from 0 to 80 GeV when $|\Delta\phi| = 0$ rad
  - from 1.5 to 0 rad when $\not\!\!E_T = 0$ GeV, and $\not\!\!E_T$ from 0 to 50 GeV when $|\Delta\phi| = 0$ rad
  - from 2.5 to π rad when $\not\!\!E_T = 0$ GeV, and $\not\!\!E_T$ from 0 to 30 GeV when $|\Delta\phi| = \pi$ rad

Orthogonal samples selection (for measuring the multijet backgrounds):

- All the same selection cuts as listed above except for the tight lepton requirements
- Electron channel — only one loose-but-not-tight electron
- Muon channel — only one loose-but-not-tight muon

Table 6.1: Numbers of events for the electron and muon channels after selection.

	Numbers of Events after Selection									
	Electron Channel					Muon Channel				
	1 jet	2 jets	3 jets	4 jets	5+ jets	1 jet	2 jets	3 jets	4 jets	5 jets
Signal MC										
tb	6,908	19,465	9,127	2,483	595	3,878	12,852	6,458	1,809	401
tqb	8,971	22,758	12,080	3,797	1,092	8,195	21,066	11,193	3,489	835
tb+tqb	15,879	42,223	21,207	6,280	1,687	12,073	33,918	17,651	5,298	1,236
Background MC										
$t\bar{t}\rightarrow ll$	7,671	29,537	26,042	12,068	5,396	5,509	24,595	21,803	9,788	3,442
$t\bar{t}\rightarrow l$+jets	522	5,659	22,477	27,319	14,298	232	3,376	16,293	22,680	8,658
$Wb\bar{b}$	26,611	13,914	9,011	3,848	1,434	27,764	14,488	9,427	3,874	1,204
$Wc\bar{c}$	21,765	13,453	7,562	2,252	591	32,712	19,047	10,141	3,051	663
Wjj	134,660	61,497	34,162	8,290	1,750	147,842	66,201	36,673	9,169	1,502
Pretag data										
Multijets	11,565	6,993	4,043	1,317	431	897	658	462	151	48
Signal data	27,370	8,220	3,075	874	223	17,816	6,432	2,590	727	173
ZeroTag data										
Multijets	11,319	6,659	3,802	1,210	390	866	604	409	128	36
Data	26,925	7,833	2,831	752	178	17,527	6,122	2,378	599	125
OneTag data										
Multijets	246	322	226	93	34	31	51	49	21	8
Signal data	445	357	207	97	35	289	287	179	100	38
TwoTags data										
Multijets	—	12	15	14	7	—	3	4	1	4
Signal data	—	30	37	22	10	—	23	32	27	10

CHAPTER 7

Bayesian Neural Networks

The identification of the signal processes over background processes is a very important task in high energy physics. In order to improve in signal/background discrimination, high energy physicists have used various multivariate statistical techniques. Nowadays, a neural network is one of the most used tools because it can handle problems with many parameters and provide good classification results.

A neural network [66] is an approach using the inductive strategy to estimate models from data. This estimation process is referred to as "learning" or "learning from data". Here, learning implies obtaining representation of data or finding patterns in the data which can be used for several purposes such as predicting future events or classification.

How learning operates can be modeled in many different ways with many different perspectives. One of the most successful approaches to describe physical processes is multivariate input-output mapping:

$$y = f(x, \theta), \tag{7.1}$$

where x corresponds to a group of input variables, y to the output or target variables, and θ to a set of unknown model parameters. In this context, learning is regarded as parameter estimation to obtain a description of the conditional distribution $p(y|x, \theta)$ on the basis of instances presented to us.

Conventional methods of learning in neural networks are interpreted as an implementation of the statistical procedure of maximum likelihood. Such approaches can suffer from several deficiencies which will be discussed later. On the other hand, The Bayesian school of statistics [67] is based on a different view of learning, in which probability is used to represent uncertainty about the relationship being learned.

In this analysis, we use Bayesian neural networks [68] to extract maximum information from the data. This chapter gives an introductory account of Bayesian methods and their application to neural networks, with the focus on underlying principles. Section 7.1 describes basic concepts of neural networks. Section 7.2 gives an outline of the theory of Bayesian statistics. Section 7.3 describes basic idea of Bayesian neural networks. Finally, details of Markov Monte Carlo and implementation of Bayesian neural networks are given in Section 7.4.

7.1 Basic Concepts of Neural Networks

The most commonly used neural networks are *multilayer perceptron* networks, also known as "feedforward" or "backpropagation" networks. This class of networks consists of a number of simple processing units, organized in layers. Every unit in a layer is connected to all the units in the previous layer. These connections are not all equal, each connection may have a different weight. The weights on these connections encode the knowledge of a network. Often the units in a neural network are also called nodes. In this network, the information moves in one direction, from the input nodes, through the hidden nodes and to the output nodes.

The network is a non-linear function $y(x; w)$, where x are input variables and w are parameters generally referred to as weights,

$$y(x; w) = \frac{1}{1 + e^{-f(x;w)}}, \tag{7.2}$$

where

$$f(x; w) = b + \sum_j v_j h_j(x), \tag{7.3}$$

$$h_j(x) = \tanh(a_j + \sum_i u_{ij} x_i). \tag{7.4}$$

Here, i runs over the inputs $x_1 ... x_I$ and j runs over the hidden units. u_{ij} is the weight on the connection from input unit i to hidden unit j; similarly, v_j is the weight on the connection from hidden unit j to output unit. The a_j and b are the biases of the hidden and output units, respectively. These weights and biases are the parameters of the network represented by w.

The output of the function y can be interpreted as the probability of input variable x that belongs to a given class. In the process of learning or training neural networks, one determines the values for w using a set of examples, referred to as training data, and build decision rules to assign class membership for inputs. The training data provide the relationship between input x and a target t. Thus, the neural networks learn a model of the relationship between x and t, and give an output y close to the target value t.

Typically an objective function or error function, defined as a function of w, is used to measure how well the network, with its weights set to w, solves the task. The error function is a sum of terms, for each input/target pair $\{x, t\}$, measuring how close the output y is to the target t. The training is an exercise in *function minimization* - *i.e.*, adjusting w in such a way as to find a w that minimizes the error function. Many function minimization algorithms make use not only of the error function, but also its *gradient* with respect to the parameters w. For neural networks the **backpropagation** algorithm [66] efficiently evaluates the gradient of the output y with respect to the parameters w, and thence the gradient of the error function with respect to w.

Eq. (7.5) shows the sum-of-squares error function which is one of the simplest and most commonly used objective functions,

$$E(w) = \frac{1}{2} \sum_{i=1}^{N} |y(x_n; w) - t_n|^2. \tag{7.5}$$

Minimizing this error function with respect to w leads to an estimate w^* which can be used to make predictions for new values of x by evaluating $y(x; w^*)$.

7.1.1 Limitation of (Conventional) Neural Networks

A well-known problem with error function minimization is that complex and flexible models can over-fit the training data, leading to poor generalization. Indeed, when the number of parameters equals the number of data points, the least squares solution for a model can achieve a perfect fit to the training data while having very poor performance on new data. This behavior is characterized by values of the parameters w that have large positive and negative values finely tuned to the individual noisy data. The corresponding function $y(x; w)$ typically exhibits strong oscillations as a function of x. While over-fitting can be avoided by limiting the complexity of the model, this too can lead to poor generalization if the model

is insufficiently flexible to capture the underlying behavior of the data. However, we often have to work with data of limited size and yet we wish to be able to use flexible models with many adjustable parameters. We shall see that the phenomenon of over-fitting is a pathological property of point estimation, and that by adopting a Bayesian viewpoint we can apply complex models to small data which minimize the problems of over-fitting.

7.2 Bayes' Theorem

We are quite used to the idea of dealing with uncertainty in our lives. For instance, we might believe that it is unlikely to rain tomorrow if the last few days have been sunny. However, if we discover that a cold front is about to arrive, we might revise our views and decide that is is in fact quite likely to rain. Here we are discussing subjective beliefs, and the way they are modified when we obtain more information. We might seek to put such reasoning on a more formal footing, and to quantify our uncertainty by encoding the degrees of belief as real numbers. Ref. [69] shows that, provided we impose some simple consistency requirements, then these numbers obey the rules of conventional probability theory. In other words, if we use a value of 1 to denote complete certainty that an event will occur, and 0 to denote complete certainty that the event will not occur, with intermediate values representing corresponding degrees of belief, then these real values behave exactly like probabilities.

Once our beliefs have been represented as probabilities they can be manipulated using two simple rules. Consider a pair of random variables A and B each of which can take on a number of discrete values. We denote by P(a,b) the joint probability that A = a and B = b. Using the product rule this joint probability can be expressed in the form

$$P(a, b) = P(b|a)P(a). \tag{7.6}$$

Here $P(b|a)$ denotes the conditional probability, in other words the probability that B = b given that A = a. We can similarly consider a conditional probability of the form $P(a|b)$. The quantity P(a) in Eq. (7.6) denotes the marginal probability, in other words the probability that A = a irrespective of the value of B. The second relation between probabilities that we need to consider is the sum rule given by

$$\sum_b P(a, b) = P(a). \tag{7.7}$$

where the sum is over all possible values of b. From the product rule we obtain the following relation

$$P(a|b) = \frac{P(b|a)P(a)}{P(b)}, \tag{7.8}$$

which is known as Bayes' theorem. Using the sum rule, we see that the denominator in Eq. (7.8) is given by

$$P(b) = \sum_a P(b|a)P(a), \tag{7.9}$$

and plays the role of a normalizing factor, ensuring that the probabilities on the left hand side of Eq. (7.8) sum to one. For continuous rather than discrete variables, the probabilities are replaced by probability densities, and summations are replaced by integrations.

We can consider $P(a)$ to be the prior probability of A = a before we observe the value of B, and $P(a|b)$ to be the corresponding *posterior probability* after we have observed the value of B. Posterior probabilities play a central role in pattern recognition, and Bayes' theorem allows them to be re-expressed in terms of quantities which may be more easily calculated.

As we shall see, we can treat the problem of learning in neural networks from a Bayesian perspective simply by application of the above rules of probability. This leads to a unique formalism which in principle is simple to apply, and which can lead to some very powerful results. We shall also see, however, that the application of Bayesian inference to realistic problems presents many difficulties which require careful analytical approximations or sophisticated numerical approaches to resolve.

7.3 Bayesian Neural Networks

Conventional neural network training procedures adjust the weights and biases in the network so as to minimize a measure of error on the training data, most commonly, the sum of the squared differences between the network outputs and the targets as described in the previous section. Minimization of this error measure is equivalent to maximum likelihood estimation.

Finding the weights and biases that minimize the chosen error function is commonly done using some gradient-based optimization method, such as *backpropagation*. There are typically many local minima, but good solutions are often found despite this.

In the Bayesian approach to neural network learning, the weights and biases, w, in neural networks are learned (trained) based on a set of training data, $(x^{(1)}, t^{(1)}), ..., (x^{(n)}, t^{(n)})$, giving examples of inputs, $x^{(i)}$, and associated targets, $t^{(i)}$. The result of Bayesian learning is a probability distribution over model parameters that expresses our beliefs regarding how likely the different parameter values are given the training data. To start the process of Bayesian learning, we must define a *prior distribution*, $P(w)$, for the parameters, that expresses our initial beliefs about their values, before any data has arrived. When we observe data, this prior distribution is updated to a *posterior distribution*;

$$P(w|(x^{(1)}, t^{(1)}), ..., (x^{(n)}, t^{(n)})) = \frac{P((x^{(1)}, t^{(1)}), ..., (x^{(n)}, t^{(n)})|w)P(w)}{P((x^{(1)}, t^{(1)}), ..., (x^{(n)}, t^{(n)}))}. \tag{7.10}$$

The posterior distribution combines the likelihood function, which contains the information about w derived from observation, with the prior, which contains the information about w derived from our background knowledge. The introduction of a prior is a crucial step that allows us to go from a likelihood function to a probability distribution, and thereby allows learning to be performed using the apparatus of probability theory. The prior is also a common focus for criticism of the Bayesian approach, as some people view the choice of a prior as being arbitrary.

One objective of Bayesian neural network (BNN) learning is to find the predictive distribution for the target values in new data, $x^{(n+1)}$, given the inputs for that data, and the inputs and targets in the training data. The predictive distribution is defined by

$$\begin{aligned} &P(t^{(n+1)}|x^{(n+1)}, (x^{(1)}, t^{(1)}), ..., (x^{(n)}, t^{(n)})) \\ &\quad = \int P(t^{(n+1)}|x^{(n+1)}, w)P(w|(x^{(1)}, t^{(1)}), ..., (x^{(n)}, t^{(n)}))dw. \end{aligned} \tag{7.11}$$

However, in our analysis we define Bayesian neural networks by the function

$$y(x) = \int y(x; w)P(w|(x^{(1)}, t^{(1)}), ..., (x^{(n)}, t^{(n)}))dw. \tag{7.12}$$

Since the distribution of the inputs is not being modeled, the likelihood is given by

$$P(t|x, w) = \prod_{i=1}^{n} P(t^{(i)}|x^{(i)}, w), \tag{7.13}$$

where for classification models, the likelihood function is taken to be

$$P(t|x,w) = \prod_{i=1}^{n} y(x^{(i)},w)^{t^{(i)}}[1-y(x^{(i)},w)]^{1-t^{(i)}}. \tag{7.14}$$

The neural network output y can be interpreted as probability that the target is 1. Similarly, probability for t=0 is 1-y. It is assumed that the training events are independent.

7.3.1 Priors

We must define the prior for the model parameters, which we express by a *hierarchical model* with a set of common *hyperparameters*, say γ.

If the w_k are independent given γ, we will have

$$P(w) = P(w_1,...,w_p) = \int P(\gamma)\prod_{i=1}^{p} P(w_i|\gamma)d\gamma. \tag{7.15}$$

The prior density for the parameters, w, is written as $P(w|\gamma)$, where the prior density for the hyperparameters themselves is $P(\gamma)$. With this hierarchical prior, the posterior density is given by

$$\begin{aligned} &P(w,\gamma|(x^{(1)},t^{(1)}),...,(x^{(n)},t^{(n)})) \\ &\quad \propto P(\gamma)P(w|\gamma)\prod_{i=1}^{n} P(t^{(i)}|x^{(i)},w,\gamma). \end{aligned} \tag{7.16}$$

Consider the network function given in Eqs. (7.2) $\sim$ (7.4). Let the parameters in a particular group of parameters be $u_1,...,u_k$ and let them have Gaussian distributions with mean zero and standard deviation σ_u. It is convenient to represent this standard deviation in terms of the corresponding "precision", defined to be $\tau_u = \sigma_u^{-2}$. We take the distribution for the parameters in the group to be

$$P(u_1,...,u_k|\tau_u) = (2\pi)^{-k/2}\tau_u^{k/2}\exp(-\tau_u\sum_{i=1} u_i^2/2). \tag{7.17}$$

The precision is given a Gamma distribution with some mean, w_u, and shape parameter specified by α_u, with density

$$P(\tau_u) = \frac{(\alpha_u/2w_u)^{\alpha_u/2}}{\Gamma(\alpha_u/2)}\tau_u^{\alpha/2-1}\exp(-\tau_u\alpha_u/2w_u). \tag{7.18}$$

In the previous notation, τ_u is a component of γ. The values of w_u and α_u may be considered fixed.

The evaluation of the BNN involves an integral over weight space, which is a challenging task. The only feasible approach to perform this integral is to use Markov chain Monte Carlo methods to sample from Eq. (7.12). The following section describes the theory and application of Markov chain Monte Carlo methods.

7.4 Markov Chain Monte Carlo (MCMC)

The Monte Carlo method, which is used extensively in high energy physics, was developed to use random numbers to compute integrals. Suppose we wish to calculate a complicated integral

$$\int_a^b h(x)dx. \tag{7.19}$$

We decompose $h(x)$ into the product of a function $f(x)$ and a probability density function $p(x)$,

$$\int_a^b h(x)dx = \int_a^b f(x)p(x)dx = E[f(x)], \tag{7.20}$$

so that the integral can be expressed as an expectation of $f(x)$ with respect to the probability density $p(x)$. If we draw random numbers, $x_1, ...x_n$ from the density $p(x)$, then

$$\int_a^b f(x)p(x)dx \simeq \frac{1}{n}\sum_{i=1}^{n} f(x_i). \tag{7.21}$$

This is referred to as **Monte Carlo integration**. A Markov chain is one way to draw random numbers from a density $p(x)$.

7.4.1 Theory of Markov Chains

A Markov chain is a series of random variables, $X^{(0)}, X^{(1)}, X^{(2)}$, ..., in which the distribution of $X^{(n+1)}$ depends only on $X^{(n)}$. More formally,

$$Pr(X^{(n+1)} = s_j | X^{(0)} = s_k, ..., X^{(n)} = s_i) = Pr(X^{(n+1)} = s_j | X^{(n)} = s_i), \tag{7.22}$$

where $X^{(n)}$ denotes the value of a random variable at time step n. The range of possible X values defines the **state space**. For a Markov random variable the only information

about the past needed to predict the future is the current state of the random variable. The transition probability, $P(i,j) = P(i \rightarrow j)$, is the probability that a process in state s_i moves to state s_j in a single step,

$$P(i,j) = P(i \rightarrow j) = Pr(X^{(n+1)} = s_j | X^{(n)} = s_i). \tag{7.23}$$

Let

$$\pi_j(n) = Pr(X^{(n)} = s_j), \tag{7.24}$$

denote the probability that the chain is in state j at time n, and let $\pi(n)$ denote the vector of the state space probabilities at step n. We start the chain by specifying a starting vector $\pi(0)$. Often all the elements of $\pi(0)$ are zero except for a single element corresponding to the process starting in that particular state. As the chain progresses, the probability values get spread out over the state space.

The probability that the chain is in state s_i at time (or step) n + 1 can be formulated as sums over the probability of being in a particular state at the current step and the transition probability from that state into state s_i,

$$\begin{aligned} \pi_i(n+1) &= Pr(X^{(n+1)} = s_i), \\ &= \sum_k Pr(X^{(n+1)} = s_i | X^{(n)} = s_k) \cdot Pr(X^{(n)} = s_k), \\ &= \sum_k P(k \rightarrow i)\pi_k(n) = \sum_k P(k,i)\pi_k(n). \end{aligned} \tag{7.25}$$

Define the **probability transition matrix P** as the matrix whose i, jth element is $P(i,j)$, the probability of moving from state i to state j, $P(i \rightarrow j)$. Eq. (7.25) becomes

$$\pi(n+1) = \pi(n)\mathbf{P}. \tag{7.26}$$

Using the matrix form, we can show that [70, 71]

$$\pi(n) = \pi(0)\mathbf{P}^n. \tag{7.27}$$

Defining the n-step transition probability $p_{ij}^{(t)}$ as the probability that the process is in state j given that it started in state i t steps ago,

$$p_{ij}^{(t)} = Pr(X^{(n+t)} = s_j | X^{(n)} = s_i), \tag{7.28}$$

it immediately follows that $p_{ij}^{(t)}$ is just the ij-th element of $\mathbf{P}^t$.

Finally, a Markov chain is said to be irreducible if there exists a positive integer such that $p_{ij}^{(t)} > 0$ for all i, j. That is, all states **communicate** with each other, as one can always go from any state to any other state (although it may take more than one step). Likewise, a chain is said to be **aperiodic** when the number of steps required to move between two states (say x and y) is not required to be a multiple of some integer. Put another way, the chain is not forced into some cycle of fixed length between certain states.

An *invariant* (or *stationary*) distribution over the states of a Markov chain is one that persists forever once it is reached. More formally, the distribution given by the probabilities π^* is invariant with respect to the Markov chain with transition probabilities $\mathbf{P}$, if

$$\pi^* = \pi^* \mathbf{P}. \tag{7.29}$$

A Markov chain can have more than one invariant distribution. If $\mathbf{P}$ is the identity matrix, for example, then any distribution is invariant. A finite Markov chain always has at least one invariant distribution.

We are interested in constructing Markov chains for which the distribution we wish to sample from, given by π^*, is invariant. Often, we will use *time reversible* homogeneous Markov chains that satisfy the more restrictive condition of *detailed balance* - that if a transition occurs from a state picked according to the probabilities given by π^*, then the probability of that transition being from state j to state k is the same as the probability of it being from state k to state j. In other words, for all j,

$$P(j, k)\pi_j^* = P(k, j)\pi_k^*. \tag{7.30}$$

This also shows that π^* is an invariant distribution. However, it is possible for a distribution to be invariant without detailed balance holding.

For our purposes, it is not enough merely to find a Markov chain with respect to which the distribution we wish to sample from is invariant. We also require that the Markov chain be *ergodic* - that the probabilities at time t, $\pi(t)$, converge to this invariant distribution as $t \to \infty$, regardless of the choice of initial probabilities $\pi(0)$. An ergodic Markov chain

can have only one invariant distribution, which is also referred to as its "equilibrium" distribution. Some Markov chains "converge" not to a single distribution, but rather to a cycle of distributions. These periodic chains are not ergodic by this definition.

7.4.2 Metropolis Algorithm

One problem with applying Monte Carlo integration is in obtaining samples from some complex probability distribution $p(x)$. Attempts to solve this problem are the roots of MCMC methods. In particular, they trace to attempts by mathematical physicists to integrate very complex functions by random sampling (Metropolis and Ulam 1949, Metropolis et al. 1953, Hastings 1970), and the resulting Metropolis-Hastings algorithm [72].

Suppose our goal is to draw samples from some distribution $p(\theta)$ where $p(\theta) = f(\theta)/K$, where the normalizing constant K may not be known, and very difficult to compute. The Metropolis algorithm (Metropolis and Ulam 1949, Metropolis et al. 1953) generates a sequence of draws from this distribution as follows:

1. Start with any initial value θ_0 satisfying $f(\theta_0) > 0$.

2. Using the current θ value, sample a candidate point θ^* from some jump density $q(\theta_1; \theta_2)$, which is the probability of returning a value of θ_2 given a previous value of θ_1. This distribution is also referred to as the proposal or candidate-generating distribution. The only restriction on the jump density in the Metropolis algorithm is that it be symmetric, i.e., $q(\theta_1; \theta_2) = q(\theta_2; \theta_1)$.

3. Given the candidate point θ^*, calculate the ratio of the density at the candidate (θ^*) and current (θ_{t-1}) points,

$$\alpha = \frac{p(\theta^*)}{p(\theta_{t-1})} = \frac{f(\theta^*)}{f(\theta_{t-1})}. \tag{7.31}$$

Notice that because we are considering the ratio of $p(x)$ under two different values, the normalizing constant K cancels out.

4. If the jump increases the density ($\alpha > 1$), accept the candidate point (set $\theta_t = \theta^*$) and return to step 2. If the jump decreases the density ($\alpha < 1$), then with probability α accept the candidate point, else reject it and return to step 2.

We can summarize the Metropolis sampling as first computing

$$\alpha = min(\frac{f(\theta^*)}{f(\theta_{t-1})}, 1), \tag{7.32}$$

and then accepting a candidate point with probability α (the probability of a move). This generates a Markov chain (θ_0, ..., θ_k, ...), as the transition probabilities from θ_t to θ_{t+1} depends only on θ_t and not (θ_0, ..., θ_{t-1}). Following a sufficient $burn - in\ period$ (of, say, k steps), the chain approaches its stationary distribution and (as we will demonstrate shortly), samples from the vector (θ_{k+1}, ..., θ_{k+n}) are samples from $p(x)$.

Hastings [72] generalized the Metropolis algorithm by using an arbitrary transition probability function $q(\theta_1, \theta_2) = Pr(\theta_1 \rightarrow \theta_2)$, and setting the acceptance probability for a candidate point as

$$\alpha = min(\frac{f(\theta^*)q(\theta^*, \theta_{t-1})}{f(\theta_{t-1}q(\theta_{t-1}), \theta^*)}, 1). \tag{7.33}$$

This is the $Metropolis - Hastings\ algorithm$. Assuming that the proposal distribution is symmetric, i.e., $q(x, y) = q(y, x)$, recovers the original Metropolis algorithm.

7.4.3 Gibbs Sampling

The *Gibbs* sampler, also known as the *heatbath* algorithm, is conceptually the simplest of the Markov chain sampling methods, but has come into prominence only recently, with the work of Geman and Geman [73] and Gelfand and Smith [74]. It is widely applicable to problems where the variables take on values from a small finite set, or have conditional distributions of a parametric form that can easily be sampled from.

Suppose we wish to sample from the joint distribution for $X = (X_1, ..., X_n)$ given by $P(x_1, ..., x_n)$, where the range of the X_i may be either continuous or discrete. The Gibbs sampler does this by repeatedly replacing each component with a value picked from its distribution conditional on the current values of all other components. This process can be seen as generating a realization of a Markov chain that is built from a set of base transition probabilities B_k, for $k = 1$, ..., n, with

$$B_k(x, x') = P(x'_k | \{x_i : i \neq k\}) \prod_{i \neq k} \delta(x_i, x'_i), \tag{7.34}$$

i.e., B_k leaves all the components except x_k unchanged, and draws a new x_k from its distribution conditional on the current values of all the other components. This is assumed to be a feasible operation. These base transitions are usually applied in sequence, as in Eq. (7.34), though at each step we could instead pick a B_k at random from some pre-specified

distribution, as in Eq. (7.34). To complete the definition of the Markov chain, we also must specify some initial distribution, p0(x), but the hope is that this choice will not be critical.

When the B_k are applied in sequence, the algorithm can be described as simulating a homogeneous Markov chain, $X^{(0)}$; $X^{(1)}$; $X^{(2)}$, ..., with transition matrix $T = B_1...B_n$. The procedure for generating $X^{(t)}$ from $X^{(t-1)}$ can be expressed as follows:

Pick $X_1^{(t)}$ from the distribution for X_1 given $x_2^{(t-1)}, x_3^{(t-1)}, ..., x_n^{(t-1)}$.
Pick $X_2^{(t)}$ from the distribution for X_2 given $x_1^{(t)}, x_3^{(t-1)}, ..., x_n^{(t-1)}$.
$\vdots$
Pick $X_i^{(t)}$ from the distribution for X_i given $x_1^{(t)}, ..., x_{i-1}^{(t)}, x_{i+1}^{(t-1)}, ..., x_n^{(t-1)}$.
$\vdots$
Pick $X_n^{(t)}$ from the distribution for X_n given $x_1^{(t)}, x_2^{(t)}, ..., x_{n-1}^{(t)}$.

Note that the new value for X_{i-1} is used immediately when picking the next value for X_i.

To show that the Gibbs sampling algorithm works, we must first verify that all the B_k leave the desired distribution invariant. Intuitively, this is clear. Since B_k leaves the components x_i for $i \neq k$ unchanged, the desired marginal distribution for these components is certainly invariant. Furthermore, the conditional distribution for x_k in the new state given the other components is defined to be that which is desired. Together, these ensure that if we started from the desired distribution, the joint distribution for all the X_i after B_k is applied must also be the desired distribution.

7.4.4 The Hybrid Monte Carlo Algorithm

The hybrid Monte Carlo algorithm [75] combines the Metropolis algorithm and sampling techniques based on dynamical simulation. The output of the algorithm is sample points drawn from a specified distribution.

The hybrid Monte Carlo algorithm is expressed in terms of sampling from the canonical distribution. However, the algorithm can be used to sample from any distribution. The canonical distribution over the *phase space* of q and p is defined to be

$$P(q, p) \propto \exp(-H(q, p)), \tag{7.35}$$

where q is a position variable which has n real-valued components; p is a momentum variable, which has corresponds one-to-one to the n real-valued components q. $H(q, p) = E(q) + K(p)$

is the *Hamiltonian* function which gives the total energy. $E(q)$ is the potential energy and $K(p)$ is the kinetic energy. In this application q will be the set of network parameters (weights and biases).

Hamiltonian dynamics will explore the whole region of phase space with a given constant value of H. However, such transitions are not sufficient to produce an ergodic Markov chain since regions with different values of H are not visited. To access regions with different values of H, p is updated using Gibbs sampling.

In practice, Hamiltonian dynamics cannot be simulated exactly, but can be approximated by some discretization using finite time steps called the *leapfrog* method [68].

$$\hat{p}_i(\tau + \frac{\epsilon}{2}) = \hat{p}_i(\tau) - \frac{\epsilon}{2}\frac{\partial E}{\partial q_i}(\hat{q}(\tau)), \tag{7.36}$$

$$\hat{q}_i(\tau + \epsilon) = \hat{q}_i(\tau) + \epsilon\frac{\hat{p}_i(\tau + \frac{\epsilon}{2})}{m_i}, \tag{7.37}$$

$$\hat{p}_i(\tau + \epsilon) = \hat{p}_i(\tau + \frac{\epsilon}{2}) - \frac{\epsilon}{2}\frac{\partial E}{\partial q_i}(\hat{q}(\tau + \epsilon)). \tag{7.38}$$

Such a leapfrog step consists of a half step for the p_i, a full step for the q_i, and another half-step for the p_i. For some time period, $\Delta\tau$, and step size, ϵ, $L = \Delta\tau/\epsilon$ steps are applied in order to reach the target time.

Given values for the leapfrog stepsize, ϵ, and the number of leapfrog steps, L, the hybrid Monte Carlo algorithm is performed as follows:

1) From the current state, $(q, p) = (\hat{q}(0), \hat{p}(0))$, L leapfrog steps with a stepsize of ϵ are applied to reach the state $(\hat{q}(\epsilon L), \hat{p}(\epsilon L))$.
2) Negate the momentum variables,
   (q^*, p^*) $(\hat{q}(\epsilon L), -\hat{p}(\epsilon L))$
3) In the Metropolis algorithm, accepting the candidate state (q^*, p^*) with probability
   $$\min(\exp(-(H(q^*, p^*) - H(q, p))), 1),$$
   and otherwise letting the new state be the same as the old.

7.4.5 Implementing Bayesian Neural Networks using MCMC

In summary, the BNN output for new data is computed using Eq. (7.12). which can be approximated by the average value of the function $y(x; w)$ over a sample of values from the posterior using the Markov chain Monte Carlo method described above.

To sample using the hybrid method, momentum variable p_i is introduced. The hybrid Monte Carlo generates a candidate state, and the Metropolis algorithm accepts or rejects this candidate based on the change in total energy, $H(q, p) = E(q) + K(p)$. After this step, the hyperparameter is updated by Gibbs sampling.

CHAPTER 8

Analysis

8.1 Introduction

This chapter describes the application of Bayesian neural networks (BNN) to the single top quark search.

In the BNN analysis we attempt to approximate a one-to-one function of

$$\Pr(S|x) = \frac{f(x|S)\,\mathrm{prior}(S)}{f(x|S)\,\mathrm{prior}(S) + f(x|B)\,\mathrm{prior}(B)}, \tag{8.1}$$

the probability that an event characterized by the variables x belongs to the signal class, S. The densities $f(x|S)$ and $f(x|B)$ are the probability density functions for the signal and background, respectively, while $\mathrm{prior}(S)$ and $\mathrm{prior}(B)$ are the corresponding class prior probabilities. A cut on the probability $\Pr(S|x)$, or weighting events by it [76], minimizes the probability to misclassify events. Since *any* one-to-one function of $\Pr(S|x)$ is equivalent to $\Pr(S|x)$, in practice, we construct an approximation to the discriminant

$$D(x) = \frac{f(x|S)}{f(x|S) + f(x|B)}, \tag{8.2}$$

built using equal numbers of signal and background events. We approximate the discriminant $D(x)$ with BNN as described in Chapter 7, in particular, Eq. (7.12).

The signal consisted of an admixture of s and t channel single top Monte Carlo (MC) events, in the Standard Model (SM) ratio, and the background an admixture of $t\bar{t}$ and W+jets MC events plus real QCD events, also mixed in the predicted proportions. We use the terms signal and background as shorthand for these admixtures.

8.1.1 Analysis Road-Map

An important step in the BNN analysis is to construct an accurate approximation to Eq. (8.2). However, before that can be done a set of variables x must be found that provide

some measure of discrimination between the signal and background. We know *a priori* what the fundamental observables are:

1. the missing transverse energy 2-vector ($\not\!\!E_T$, $\not\!\!E_{T\phi}$);

2. the lepton 4-vector $(E_T, \eta, \phi, mass)$, taking the $mass = 0$;

3. the jet 4-vector $(E_T, \eta, \phi, mass)$, assuming massless jets, and jet-type, either 1 for b-jets or 0 for non-b-jets, for jet multiplicities $n_j = 2$, 3 or 4 jets.

The number of degrees of freedom is therefore $2+3+4\times n_j$, which ranges from 13 (for 2 jets) to 21 (for 4 jets). (Since we have assumed the masses to be zero, we can drop one degree of freedom per 4-vector.) If, for each jet multiplicity bin, we were able to compute the signal and background densities $f(x|S)$ and $f(x|B)$, in terms of these degrees of freedom, we would be done. The principal virtue of this approach is that no variable selection is needed and the variables are optimal by construction. Alternatively, we can use physical intuition and insight to construct, from the fundamental degrees of freedom, variables that capture discriminating aspects of the signal vis-à-vis the background. Obviously, these derived variables contain no more information than is contained in the original degrees of freedom. The reason for using derived variables is practical: it may be easier to construct accurate approximations to $D(x)$ using the derived variables rather than the underlying degrees of freedom.

The Bayesian neural network analysis starts with the standard set of Single Top Group derived variables and culls them to useful subsets for subsequent analysis. We used the standard set of "train" Root files for training and the standard "yield" files to obtain results [77]. From this starting point, the BNN analysis reported here proceeded as follows.

1. For each of the 12 analysis channels, $(\mu, e) \times$ (1-tag, 2-tag) $\times$ $(2, 3, 4)$ $-$ jet, eliminate variables deemed poorly modeled using p-values computed using a discrepancy measure akin to the Kolmogorov-Smirnov (K-S) statistic, but which is designed for histogrammed data.

2. For each of the 12 analysis channels:

   - rank the remaining variables according to their discrimination importance using the RuleFit algorithm [78];

- select the N_{var} best variables and build a discriminant $D(x)$ using the Bayesian method, described in Chapter 7, and
- verify that the Bayesian estimate of $D(x)$ is adequate.

3. Compute the posterior density of the single top cross section using binned likelihoods formed from the BNN output distributions:
   - combine different tag and jet multiplicity channels for both lepton flavors, and optimize the bin and channel combination using a SM signal+background ensemble;
   - study the bias of the method for the single top cross section measurement using different non-SM signal+background ensembles and
   - compute the expected and observed posterior densities for each of the 12 analysis channels as well as the optimum channel combination and determine the cross sections.
4. Compute a p-value that quantifies the significance of the result using a background-only ensemble.

8.2 Variable Selection

The standard set of variables, used by the Single Top Group, are the result of an analysis of the signal and background Feynman diagrams [79, 80] and a study of single top quark production at next-to-leading order [81]. The variables fall into three categories: object kinematics, event kinematics, and variables based on angular correlations. The complete initial set of variables is shown in Table 8.1. Jets are sorted in p_{T} and index 1 refers to the leading jet in a jet category: "jet$n$" (n=1,2,3,4) corresponds to each jet in the event, "tag$n$" to b-tagged jets, "untag$n$" to non-b-tagged jets, "best$n$" to the best jet and "notbest$n$" to all but the best jet. The best jet is defined as the one for which the invariant mass $M(W, \mathrm{jet})$ is closest to $m_t = 175$ GeV.

However, for each of the 12 (lepton, tag, jet-bin)-channels only sub-sets of these variables were used as inputs to the Bayesian neural networks. Each sub-set contained variables that were both well modeled and useful as a discriminant.

8.2.1 Checking Modeling of Variables

It is clearly necessary that the variables used be well modeled. In principle, this requires that we check that the N_{var}-dimensional densities $f(x|S)$ and $f(x|B)$ are correct. Since this is exceedingly difficult to do, in practice we check their 1-dimensional projections. While this is not sufficient to guarantee that the densities are correctly modeled it is at least a necessary condition.

The standard way to eliminate a poorly modeled variable is to do a goodness-of-fit test: we compare its histogram, derived from the model, with that of the data, compute a measure of discrepancy between it and the data histogram and from the discrepancy the associated p-value. If the p-value is judged to be small enough, the hypothesis that the histograms agree is rejected. Unfortunately, there are many ways in which two histograms can differ, each suggesting a different discrepancy measure and a different p-value and it is far from obvious which, if any, measure is sufficient for the task. However, whatever the discrepancy measure it should satisfy the following requirements: 1) it yields a p-value for which smaller values indicate greater evidence *against* the hypothesis being tested and 2) Prob(p-value) = Uniform(0,1), that is, the distribution of p-values, assuming the hypothesis to be true, should be flat.

For histograms, the second condition holds only approximately if the p-values for the Kolmogorov-Smirnov (K-S) and χ^2 discrepancy measures are computed using the standard routines, available, for example, in Root. Therefore, to be safe, we chose to calculate, by direct Monte Carlo simulation, the distributions of these and two other discrepancy measures. These empirically derived distributions were then used to compute p-values for each variable.

For each variable, in each of the 12 channels, we compared the data histogram with the corresponding model histogram, comprising a sum of signal and background distributions. Variables with a K-S p-value below ~ 0.1 were rejected.

Examples of the results of these calculations are shown in Fig. The first plot in Fig. 8.1 shows the distribution of the best modeled variable, p_T(jet3), in the (electron, 1-tag, 4-jet) channel, along with 4 different p-values. The second plot shows the distribution and p-values for the least well modeled variable, $\cos(\text{jet1}, \text{alljets})_{\text{alljets}}$, in this channel, while the third plot shows the results for the variable Missing E_T whose p-value is near the K-S p-value rejection threshold.

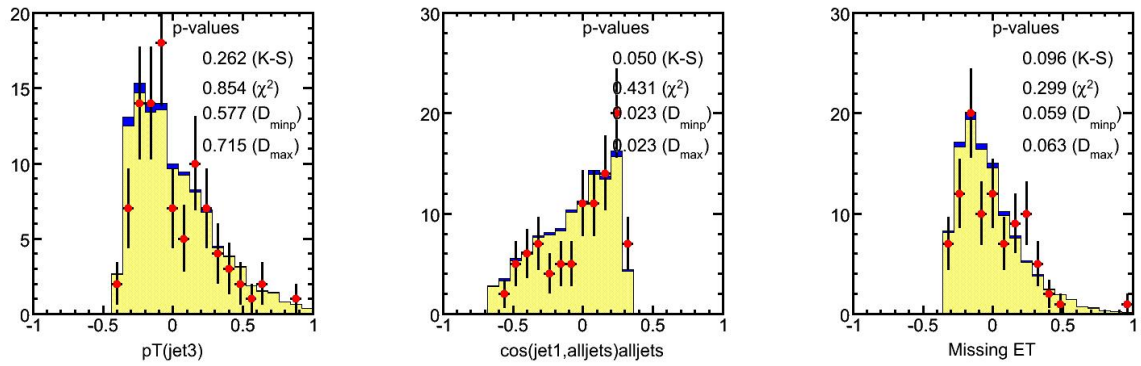


Figure 8.1: *First plot*: distributions and p-values for the best modeled variable, p_T(jet3), in the (electron, 1-tag, 4-jet) channel. *Second plot*: distributions and p-values for the worst modeled variable, $\cos(\text{jet1}, \text{alljets})_{\text{alljets}}$. *Third plot*: distributions and p-values for the variable Missing E_T, which has a K-S p-value near the rejection threshold. The 3$^{\text{rd}}$ and 4$^{\text{th}}$ p-values pertain to the discrepancy measures $D_{\text{minp}} = |C - F|(\text{min } p\text{-value})$ and $D_{\text{max}} = \max|C - F|$, respectively, where C and F, respectively, are the observed and expected counts in the associated local discrepancy, as defined in APPENDIX A. Note that each of the 4 p-values is based on a specific kind of discrepancy between two histograms. Therefore, the 4 p-values need not and indeed, do not, agree exactly.

8.2.2 Reducing Number of Variables

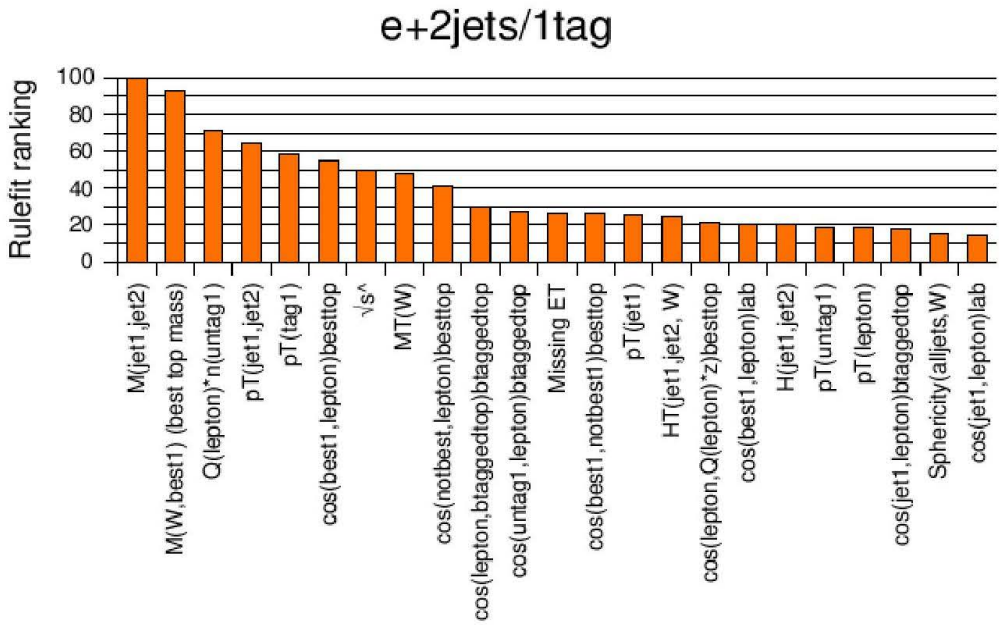
For BNN method, the computational burden increases rapidly with the dimensionality of the input data. Therefore, it pays to keep the dimensionality as low as possible. To that end, we eliminated those variables judged to provide insufficient signal/background discrimination. This was done separately for each of the 12 channels using an algorithm called RuleFit [78] to rank the importance of each variable. For each channel, we used RuleFit to create a set of trees that discriminate between the full signal and the full background. Each branch of every tree yields an "If-Then-Else" rule (that is, cuts), a weighted sum of which is created. RuleFit finds the set of coefficients in the linear sum that maximizes the discrimination between the two classes of events. Each coefficient in the linear sum is used as the weight of the corresponding rule. Roughly speaking, the importance of a variable is some measure proportional to the product of the number of times the variable appears in a rule and its weight. Variables with ranking less than $\sim$ 10 (on a scale of 0-100) were removed from further consideration.

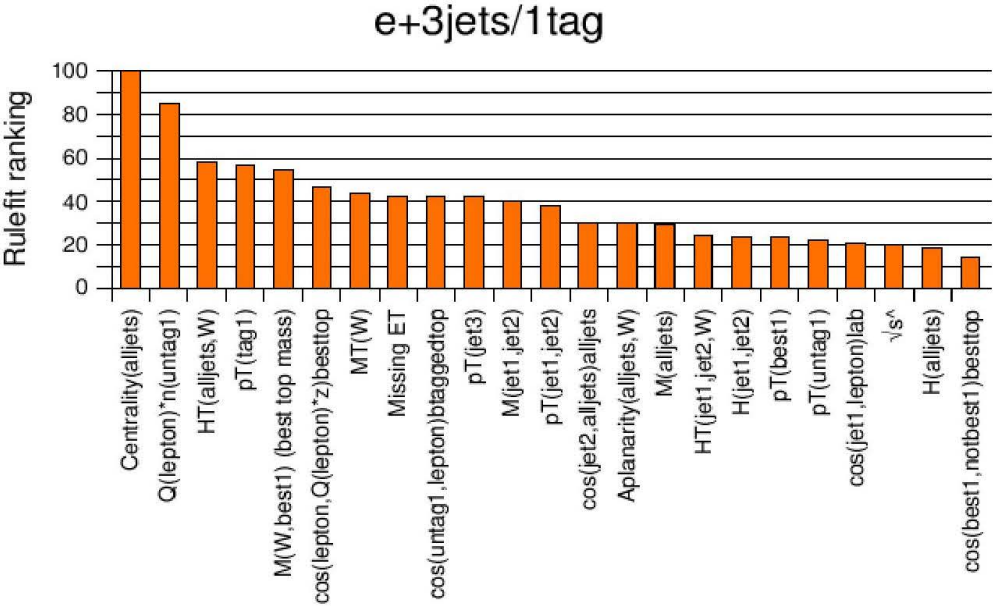
The resulting sets of well modeled variables used in this analysis, ranked according to the Rulefit importances, are shown in Figs.

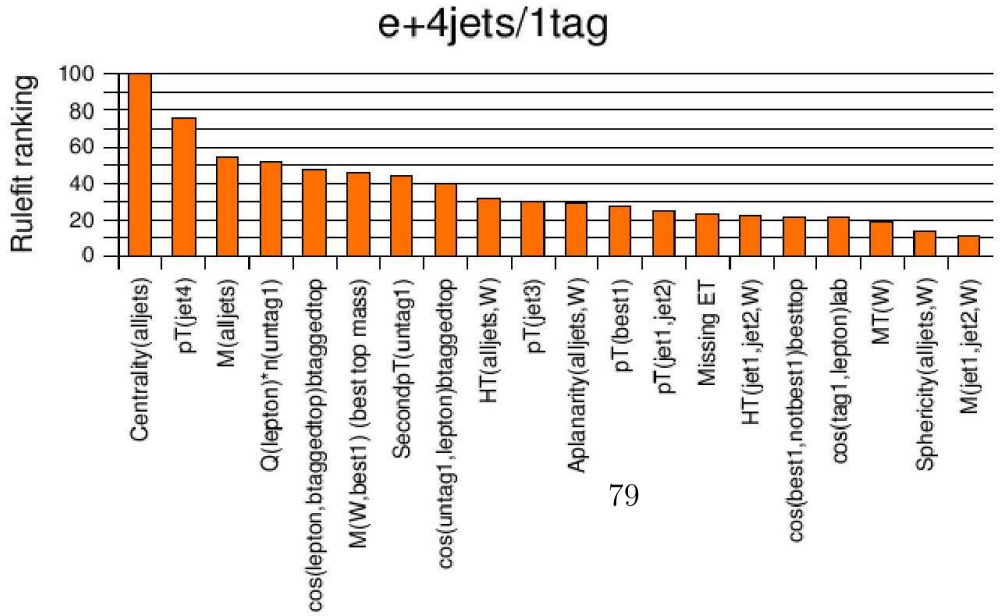
Table 8.1: Initial list of variables, in three categories: object kinematics; event kinematics and angular variables. For the angular variables, the subscript indicates the reference frame.

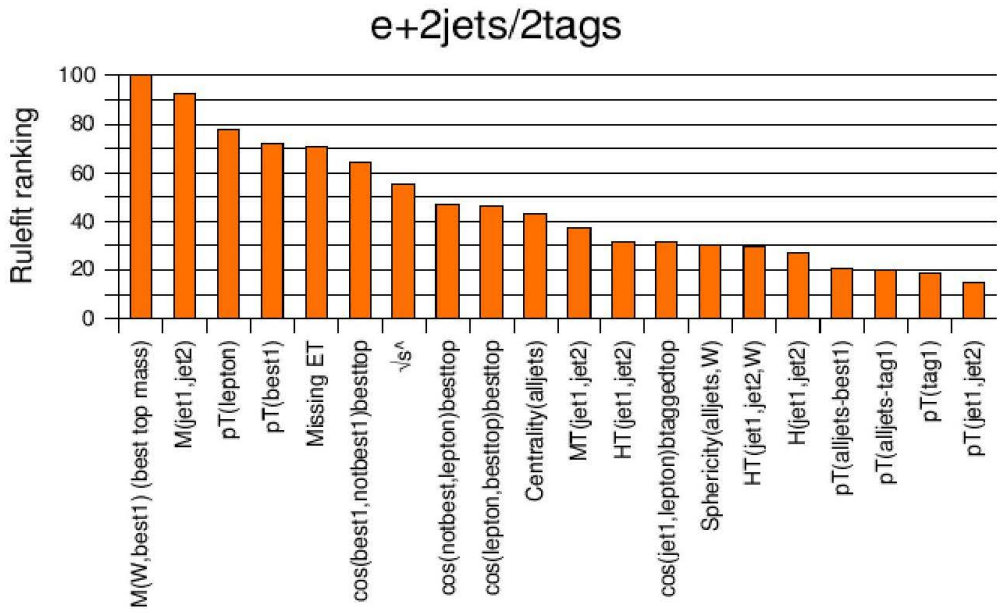
Initial list of variables

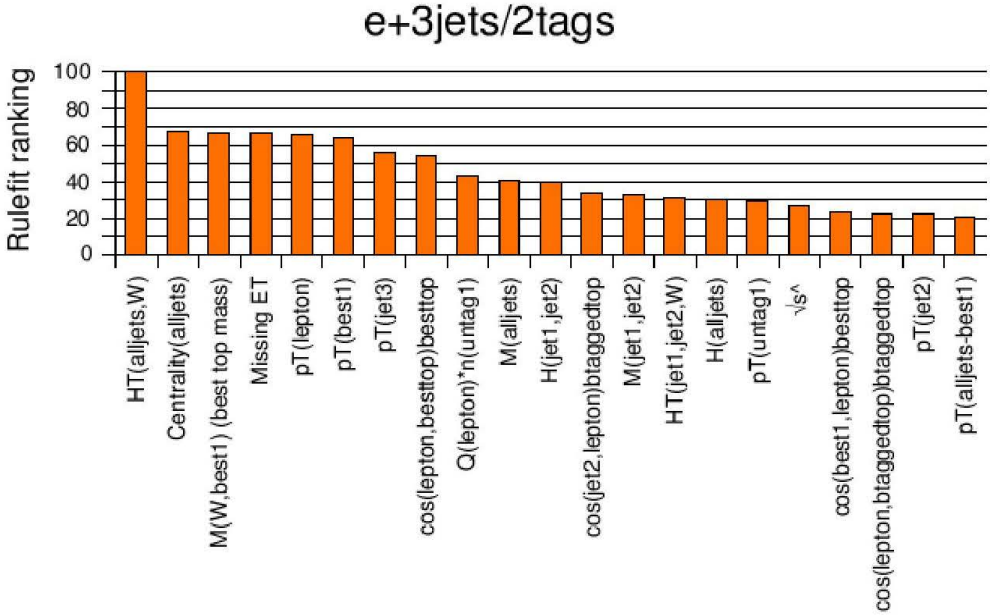
Object Kinematics	**Event Kinematics**
p_T(jet1)	Aplanarity(alljets,W)
p_T(jet2)	M(W,best1) ("best" top mass)
p_T(jet3)	M(W,tag1) ("b-tagged" top mass)
p_T(jet4)	H_T(alljets)
p_T(notbest1)	H_T(alljets−best1)
p_T(tag1)	H_T(alljets−tag1)
p_T(untag1)	H_T(alljets,W)
p_T(lepton)	H_T(jet1,jet2)
η(jet1)	H_T(jet1,jet2,W)
η(jet2)	M(alljets)
η(lepton)	M(alljets−best1)
	M(alljets−tag1)
Angular Correlations	M(jet1,jet2)
ΔR(jet1,jet2)	M(jet1,jet2,W)
cos(best1,lepton)$_{\text{besttop}}$	M_T(jet1,jet2)
cos(best1,notbest1)$_{\text{besttop}}$	M_T(W)
cos(tag1,alljets)$_{\text{alljets}}$	Missing E_T
cos(tag1,lepton)$_{\text{btaggedtop}}$	Q(lepton)×η(untag1)
cos(jet1,alljets)$_{\text{alljets}}$	$\sqrt{\hat{s}}$
cos(jet1,lepton)$_{\text{btaggedtop}}$	Sphericity(alljets,W)
cos(jet2,alljets)$_{\text{alljets}}$	Centrality(alljets)
cos(jet2,lepton)$_{\text{btaggedtop}}$	H(alljets)
cos(lepton,Q(lepton)×z)$_{\text{besttop}}$	H(jet1,jet2)
cos(lepton,besttopframe)$_{\text{besttopCMframe}}$	
cos(lepton,btaggedtopframe)$_{\text{btaggedtopCMframe}}$	
cos(notbest,alljets)$_{\text{alljets}}$	
cos(notbest,lepton)$_{\text{besttop}}$	
cos(untag1,alljets)$_{\text{alljets}}$	
cos(untag1,lepton)$_{\text{btaggedtop}}$	
cos(tag1,lepton)$_{\text{lab}}$	
cos(jet1,lepton)$_{\text{lab}}$	
cos(jet2,lepton)$_{\text{lab}}$	
ΔR(lepton,jet1)	
ΔR(lepton,jet2)	

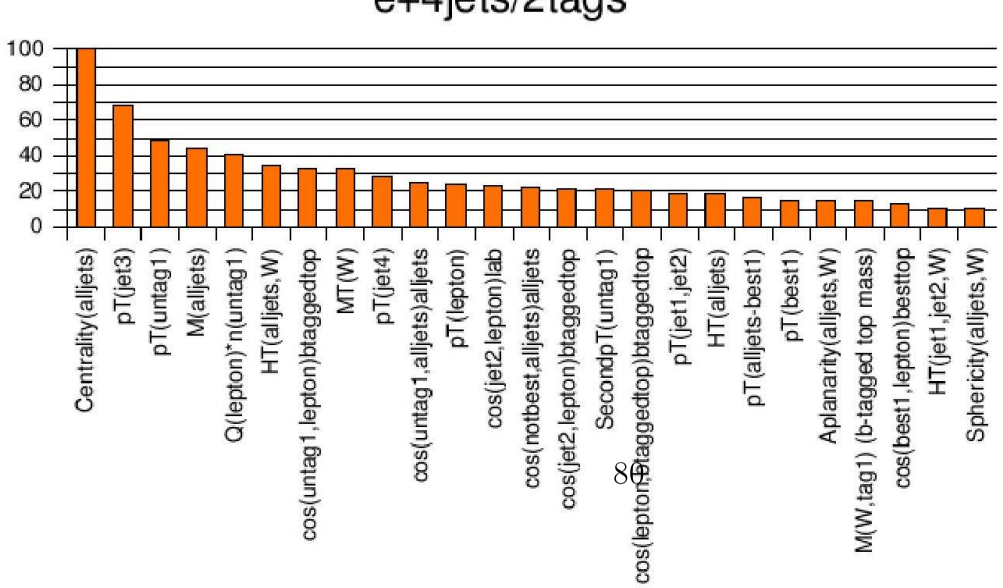

e+2jets/1tag
Rulefit ranking
100
80
60
40
20
0
M(jet1,jet2)
M(W,best1) (best top mass)
Q(lepton)*n(untag1)
pT(jet1,jet2)
pT(tag1)
cos(best1,lepton)besttop
√s^
MT(W)
cos(notbest,lepton)besttop
cos(lepton,btaggedtop)btaggedtop
cos(untag1,lepton)btaggedtop
Missing ET
cos(best1,notbest1)besttop
pT(jet1)
HT(jet1,jet2, W)
cos(lepton,Q(lepton)*z)besttop
cos(best1,lepton)lab
H(jet1,jet2)
pT(untag1)
pT(lepton)
cos(jet1,lepton)btaggedtop
Sphericity(alljets,W)
cos(jet1,lepton)lab


e+3jets/1tag
Rulefit ranking
100
80
60
40
20
0
Centrality(alljets)
Q(lepton)*n(untag1)
HT(alljets,W)
pT(tag1)
M(W,best1) (best top mass)
cos(lepton,Q(lepton)*z)besttop
MT(W)
Missing ET
cos(untag1,lepton)btaggedtop
pT(jet3)
M(jet1,jet2)
pT(jet1,jet2)
cos(jet2,alljets)alljets
Aplanarity(alljets,W)
M(alljets)
HT(jet1,jet2,W)
H(jet1,jet2)
pT(best1)
pT(untag1)
cos(jet1,lepton)lab
√s^
H(alljets)
cos(best1,notbest1)besttop

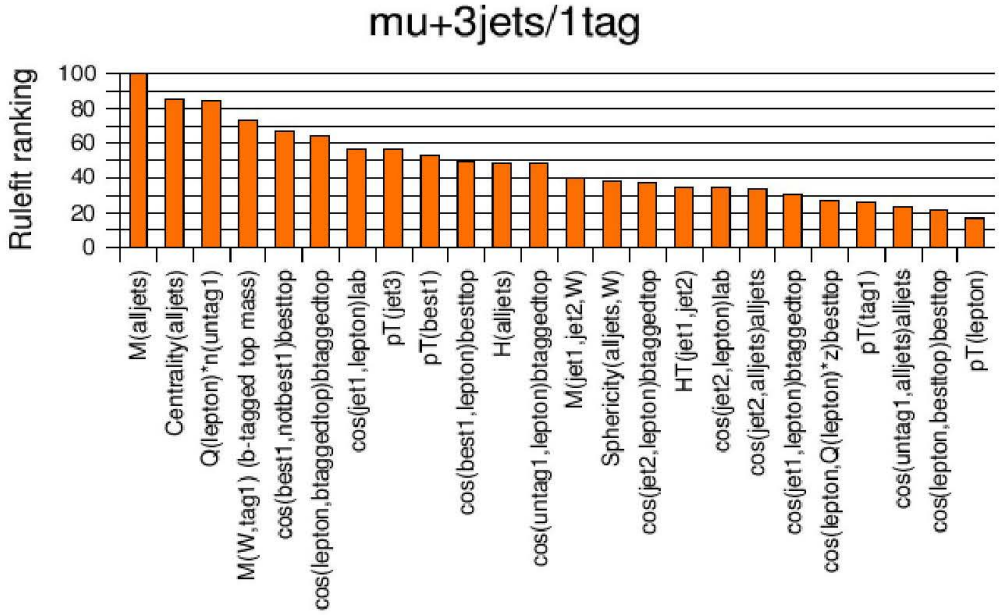

e+4jets/1tag
Rulefit ranking
100
80
60
40
20
0
Centrality(alljets)
pT(jet4)
M(alljets)
Q(lepton)*n(untag1)
cos(lepton,btaggedtop)btaggedtop
M(W,best1) (best top mass)
SecondpT(untag1)
cos(untag1,lepton)btaggedtop
HT(alljets,W)
pT(jet3)
Aplanarity(alljets,W)
pT(best1)
pT(jet1,jet2)
Missing ET
HT(jet1,jet2,W)
cos(best1,notbest1)besttop
cos(tag1,lepton)lab
MT(W)
Sphericity(alljets,W)
M(jet1,jet2,W)

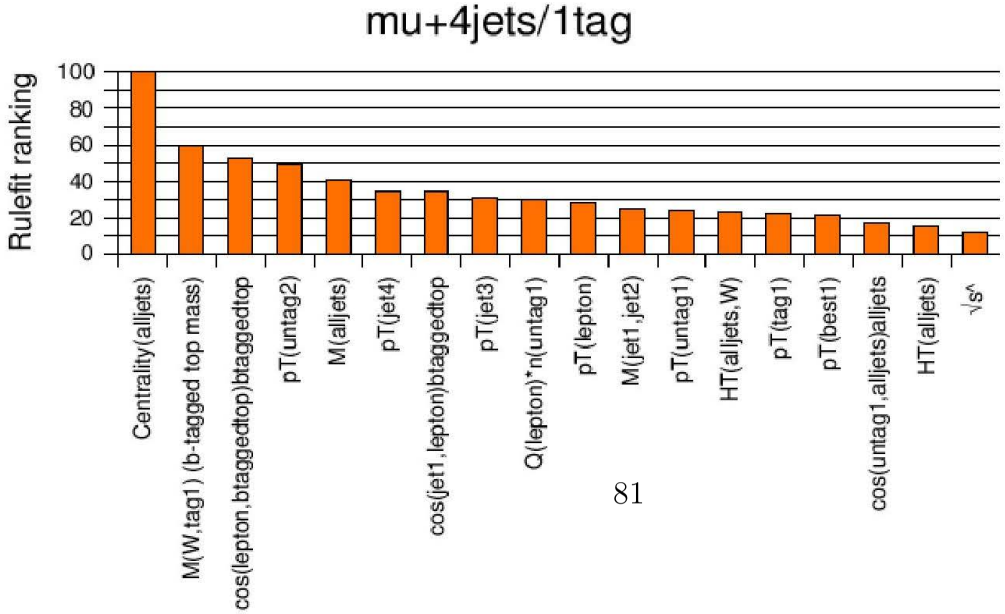

e+2jets/2tags
Rulefit ranking
100
80
60
40
20
0
M(W,best1) (best top mass)
M(jet1,jet2)
pT(lepton)
pT(best1)
Missing ET
cos(best1,notbest1)besttop
√s^
cos(notbest,lepton)besttop
cos(lepton,besttop)besttop
Centrality(alljets)
MT(jet1,jet2)
HT(jet1,jet2)
cos(jet1,lepton)btaggedtop
Sphericity(alljets,W)
HT(jet1,jet2,W)
H(jet1,jet2)
pT(alljets-best1)
pT(alljets-tag1)
pT(tag1)
pT(jet1,jet2)

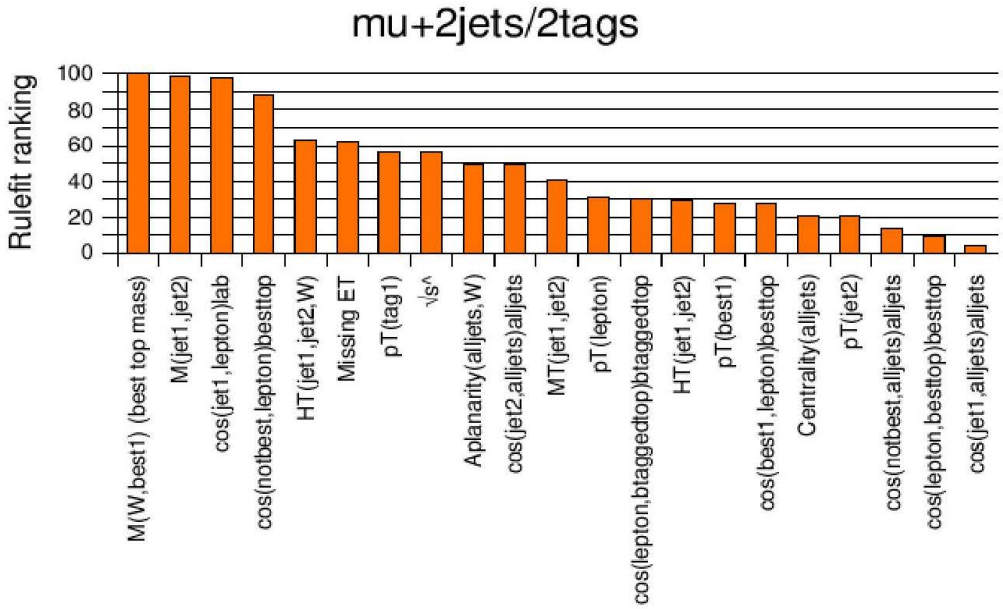

e+3jets/2tags
Rulefit ranking
100
80
60
40
20
0
HT(alljets,W)
Centrality(alljets)
M(W,best1) (best top mass)
Missing ET
pT(lepton)
pT(best1)
pT(jet3)
cos(lepton,besttop)besttop
Q(lepton)*n(untag1)
M(alljets)
H(jet1,jet2)
cos(jet2,lepton)btaggedtop
M(jet1,jet2)
HT(jet1,jet2,W)
H(alljets)
pT(untag1)
√s^
cos(best1,lepton)besttop
cos(lepton,btaggedtop)btaggedtop
pT(jet2)
pT(alljets-best1)

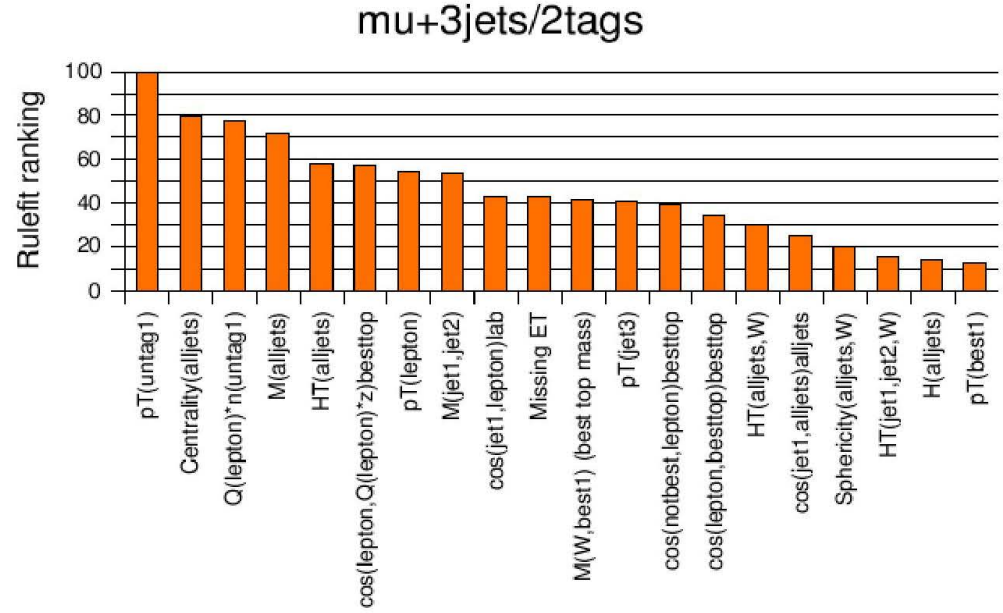

e+4jets/2tags
100
80
60
40
20
0
Centrality(alljets)
pT(jet3)
pT(untag1)
M(alljets)
Q(lepton)*n(untag1)
HT(alljets,W)
cos(untag1,lepton)btaggedtop
MT(W)
pT(jet4)
cos(untag1,alljets)alljets
pT(lepton)
cos(jet2,lepton)lab
cos(notbest,alljets)alljets
cos(jet2,lepton)btaggedtop
SecondpT(untag1)
cos(lepton,btaggedtop)btaggedtop
pT(jet1,jet2)
HT(alljets)
pT(alljets-best1)
pT(best1)
Aplanarity(alljets,W)
M(W,tag1) (b-tagged top mass)
cos(best1,lepton)besttop
HT(jet1,jet2,W)
Sphericity(alljets,W)

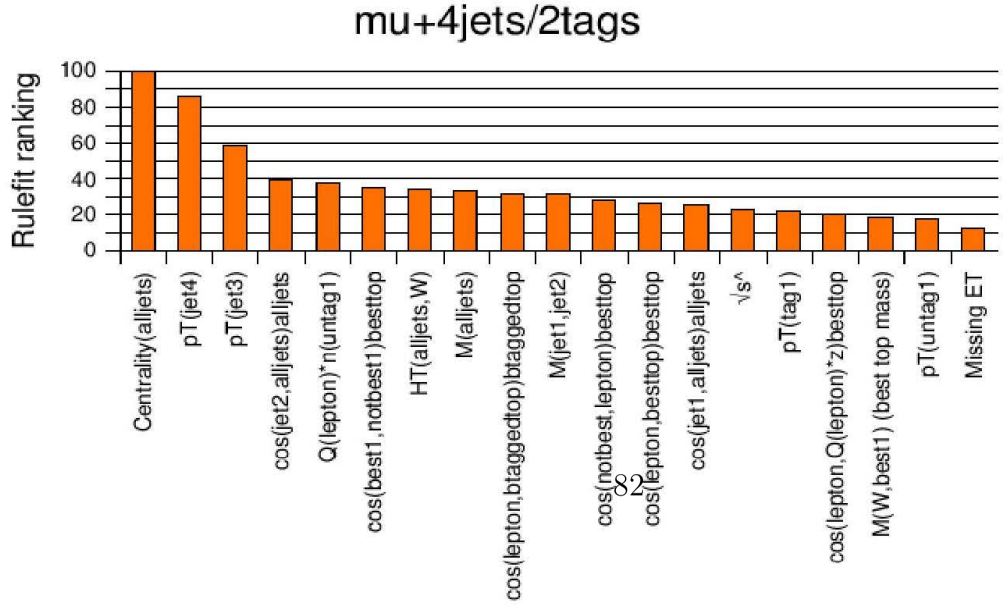

mu+2jets/1tag
100
80
60
40
20
0
M(W,tag1) (b-tagged top mass)
M(jet1,jet2)
H(jet1,jet2)
pT(best1)
Q(lepton)*n(untag1)
√s^
cos(untag1,alljets)alljets
cos(lepton,Q(lepton)*z)besttop
pT(tag1)
cos(tag1,lepton)btaggedtop
HT(jet1,jet2,W)
cos(best1,notbest1)besttop
cos(jet2,alljets)alljets
Centrality(alljets)
cos(tag1,lepton)lab
MT(jet1,jet2)
cos(jet1,lepton)lab
pT(jet2)
Sphericity(alljets,W)
MT(W)
pT(jet1,jet2)
pT(lepton)


mu+3jets/1tag
Rulefit ranking
100
80
60
40
20
0
M(alljets)
Centrality(alljets)
Q(lepton)*n(untag1)
M(W,tag1) (b-tagged top mass)
cos(best1,notbest1)besttop
cos(lepton,btaggedtop)btaggedtop
cos(jet1,lepton)lab
pT(jet3)
pT(best1)
cos(best1,lepton)besttop
H(alljets)
cos(untag1,lepton)btaggedtop
M(jet1,jet2,W)
Sphericity(alljets,W)
cos(jet2,lepton)btaggedtop
HT(jet1,jet2)
cos(jet2,lepton)lab
cos(jet2,alljets)alljets
cos(jet1,lepton)btaggedtop
cos(lepton,Q(lepton)*z)besttop
pT(tag1)
cos(untag1,alljets)alljets
cos(lepton,besttop)besttop
pT(lepton)


mu+4jets/1tag
Rulefit ranking
100
80
60
40
20
0
Centrality(alljets)
M(W,tag1) (b-tagged top mass)
cos(lepton,btaggedtop)btaggedtop
pT(untag2)
M(alljets)
pT(jet4)
cos(jet1,lepton)btaggedtop
pT(jet3)
Q(lepton)*n(untag1)
pT(lepton)
M(jet1,jet2)
pT(untag1)
HT(alljets,W)
pT(tag1)
pT(best1)
cos(untag1,alljets)alljets
HT(alljets)
√s^


mu+2jets/2tags
Rulefit ranking
100
80
60
40
20
0
M(W,best1) (best top mass)
M(jet1,jet2)
cos(jet1,lepton)lab
cos(notbest,lepton)besttop
HT(jet1,jet2,W)
Missing ET
pT(tag1)
√s^
Aplanarity(alljets,W)
cos(jet2,alljets)alljets
MT(jet1,jet2)
pT(lepton)
cos(lepton,btaggedtop)btaggedtop
HT(jet1,jet2)
pT(best1)
cos(best1,lepton)besttop
Centrality(alljets)
pT(jet2)
cos(notbest,alljets)alljets
cos(lepton,besttop)besttop
cos(jet1,alljets)alljets


mu+3jets/2tags
Rulefit ranking
100
80
60
40
20
0
pT(untag1)
Centrality(alljets)
Q(lepton)*n(untag1)
M(alljets)
HT(alljets)
cos(lepton,Q(lepton)*z)besttop
pT(lepton)
M(jet1,jet2)
cos(jet1,lepton)lab
Missing ET
M(W,best1) (best top mass)
pT(jet3)
cos(notbest,lepton)besttop
cos(lepton,besttop)besttop
HT(alljets,W)
cos(jet1,alljets)alljets
Sphericity(alljets,W)
HT(jet1,jet2,W)
H(alljets)
pT(best1)


mu+4jets/2tags
Rulefit ranking
100
80
60
40
20
0
Centrality(alljets)
pT(jet4)
pT(jet3)
cos(jet2,alljets)alljets
Q(lepton)*n(untag1)
cos(best1,notbest1)besttop
HT(alljets,W)
M(alljets)
cos(lepton,btaggedtop)btaggedtop
M(jet1,jet2)
cos(notbest,lepton)besttop
cos(lepton,besttop)besttop
cos(jet1,alljets)alljets
√s^
pT(tag1)
cos(lepton,Q(lepton)*z)besttop
M(W,best1) (best top mass)
pT(untag1)
Missing ET

8.3 BNN Training and Verification

8.3.1 Training

For each channel, the structure of each neural network is fixed: each contains $N_{\rm var}$ inputs, with $N_{\rm var} \sim 20$, and $H = 20$ hidden nodes. The numbers $N_{\rm var}$ and H define a $1+(N_{\rm var}+2)H$-dimensional parameter space each point w of which corresponds to an instance of a neural network function. Using a training set $T = (x^{(1)}, t^{(1)}), ..., (x^{(n)}, t^{(n)})$, consisting of an admixture of 10,000 $s+t$ signal events and 10,000 QCD, W+jets and $t\bar{t}$ background events, mixed in the correct proportion, we constructed a posterior density $p(w|T)$ over the network parameter space (see Chapter 7). A random sample of $K = 100$ networks was drawn from the posterior density using the Markov Chain Monte Carlo technique described in Chapter 7. The samples used are approximation of the discriminant, Eq. (8.2), via a Bayesian neural network, $y(x)$, defined by

$$\begin{aligned} y(x) &= \int y(x;w)\, p(w|T)\, dw, \\ &\approx \frac{1}{K}\sum_k y(x;w_k). \end{aligned} \tag{8.3}$$

In a previous BNN analysis, we allowed precisions of the prior density for each network parameter to adapt to the noise level in the training data. However, we found that when the training data are very noisy, the prior densities can become excessively broad and the MCMC sampling (see Chapter 7) yields networks with parameters having large values that scatter over a large range. This leads to discriminant distributions that are excessively jagged and therefore bad approximations to the discriminant, Eq. (8.2).

We therefore chose to limit the average size of network parameter values and thereby favor smoother approximations to Eq. (8.2). For each of the 12 channels, we trained a single neural network as implemented in the JETNET program [82], with $N_{\rm var}$ inputs and H hidden nodes, using back-propagation [66]. As explained in Chapter 7 that algorithm typically finds a local maximum in the likelihood function in the neighborhood of the origin of the parameter space. We used averages of the parameter values of that network as the standard deviations of the prior densities that enter the posterior density $p(w|T)$. The details are provided in APPENDIX B. This choice of prior densities prevented the network parameters from being driven to large values because of excessive noise in the training sample. Consequently, the

approximations to the discriminant $D(x)$ were much smoother, yielding a much improved BNN performance.

8.3.2 Verification

As noted above, the $K = 100$ networks are sampled from the posterior density $p(w|T)$ using the Markov chain method. It is therefore necessary to check for the convergence of the Markov chain as discussed in Ref. [83]. For each channel, we created a set of plots shown in the top rows in Figs. 8.6-8.17. For each channel, we chose different variables from each of the three categories: object kinematics, event kinematics, and angular correlations. If the BNN function models the discriminant function $D(x)$ accurately then one would expect the sum of the BNN weighted signal and background densities to reproduce the unweighted signal density. This is what we found. The good agreement seen between the signal distribution, $g(x)$, (shown by the black dots) and the sum of the BNN-weighted signal and background distributions, $f_n(x|S) + f_n(x|B)$, (the red histogram) suggests that the Markov chains had converged. The bottom rows of the same figures show the corresponding BNN outputs normalized to unity, and plots of $S/\sqrt{B}$, $S/\sqrt{S+B}$, and signal efficiency (ϵ_S) versus background efficiency (ϵ_B). The spikes in the BNN outputs are due to the low statistics in the fake-lepton sample. In general, all these plots indicate that the BNNs are well-behaved.

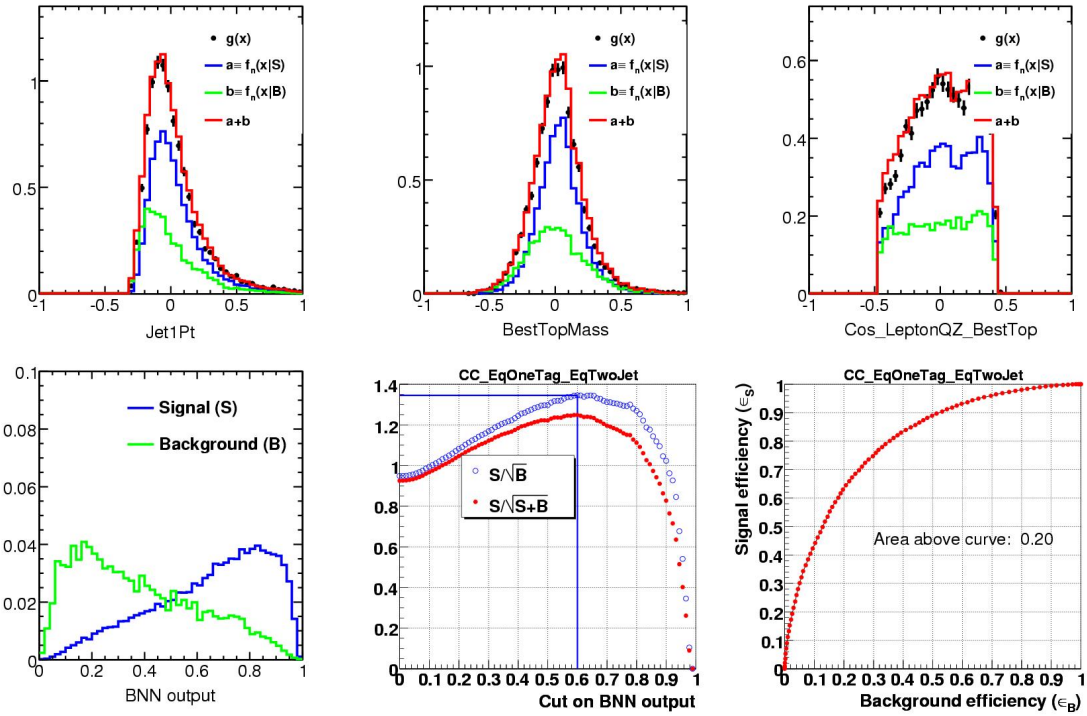


Figure 8.6: Rows: top = verification plots, bottom = BNN outputs normalized to unity (green: background, blue: signal), and the $S/\sqrt{B}$, $S/\sqrt{S+B}$, and signal efficiency (ϵ_S) versus background efficiency (ϵ_B) curves, for the electron, =1 tag, =2 jet channel.

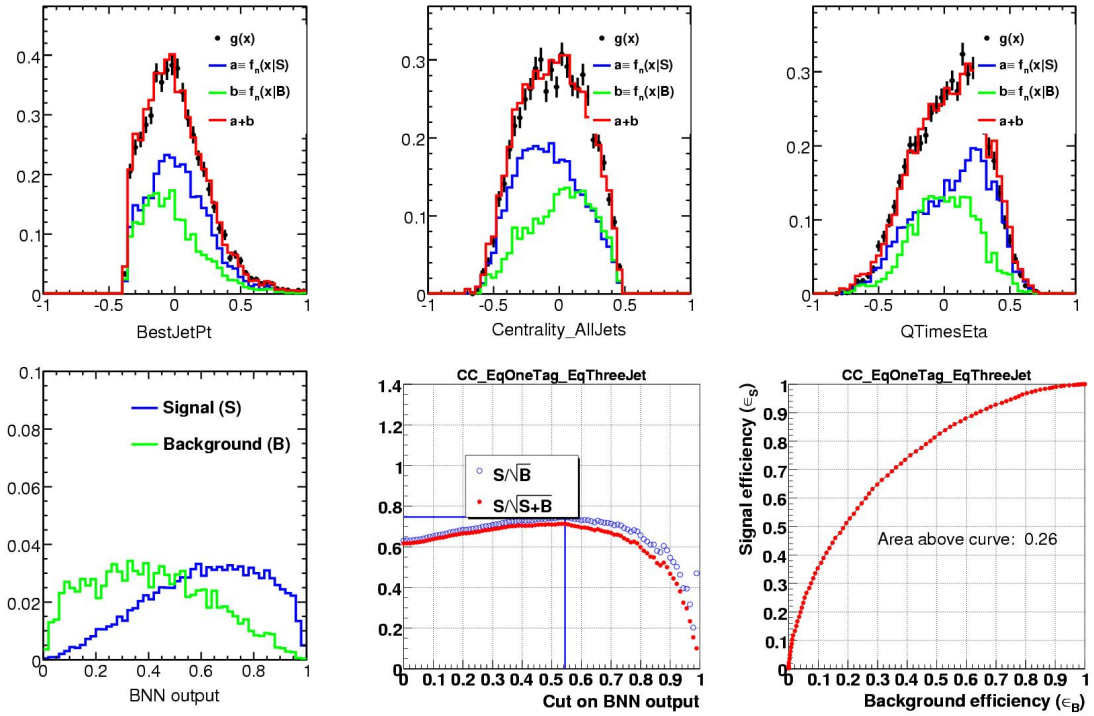


Figure 8.7: Rows: top = verification plots, bottom = BNN outputs normalized to unity (green: background, blue: signal), and the $S/\sqrt{B}$, $S/\sqrt{S+B}$, and signal efficiency (ϵ_S) versus background efficiency (ϵ_B) curves, for the electron, =1 tag, =3jet channel.

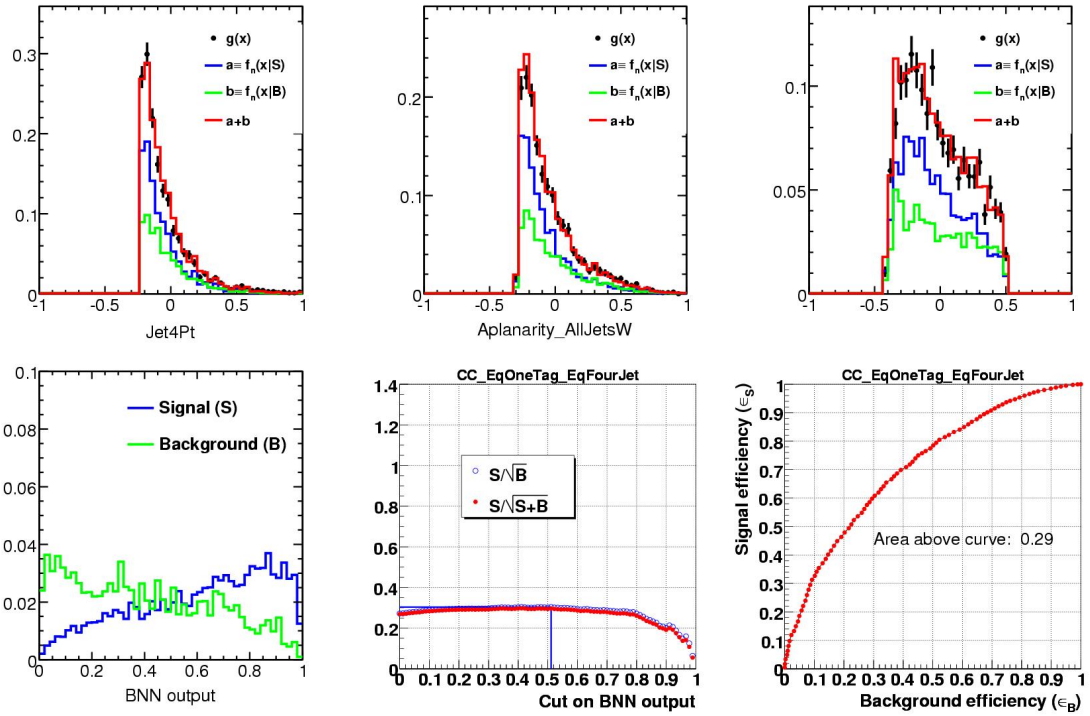


Figure 8.8: Rows: top = verification plots, bottom = BNN outputs normalized to unity (green: background, blue: signal), and the $S/\sqrt{B}$, $S/\sqrt{S+B}$, and signal efficiency (ϵ_S) versus background efficiency (ϵ_B) curves, for the electron, =1 tag, =4jet channel.

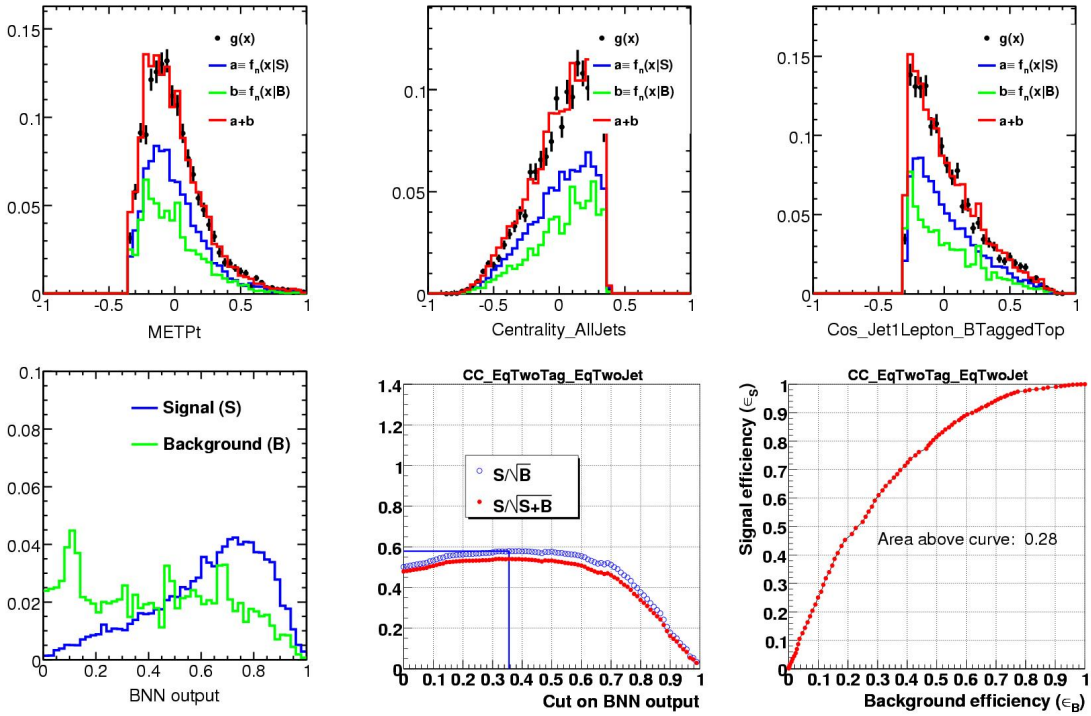


Figure 8.9: Rows: top = verification plots, bottom = BNN outputs normalized to unity (green: background, blue: signal), and the $S/\sqrt{B}$, $S/\sqrt{S+B}$, and signal efficiency (ϵ_S) versus background efficiency (ϵ_B) curves, for the electron, =2 tag, =2jet channel.

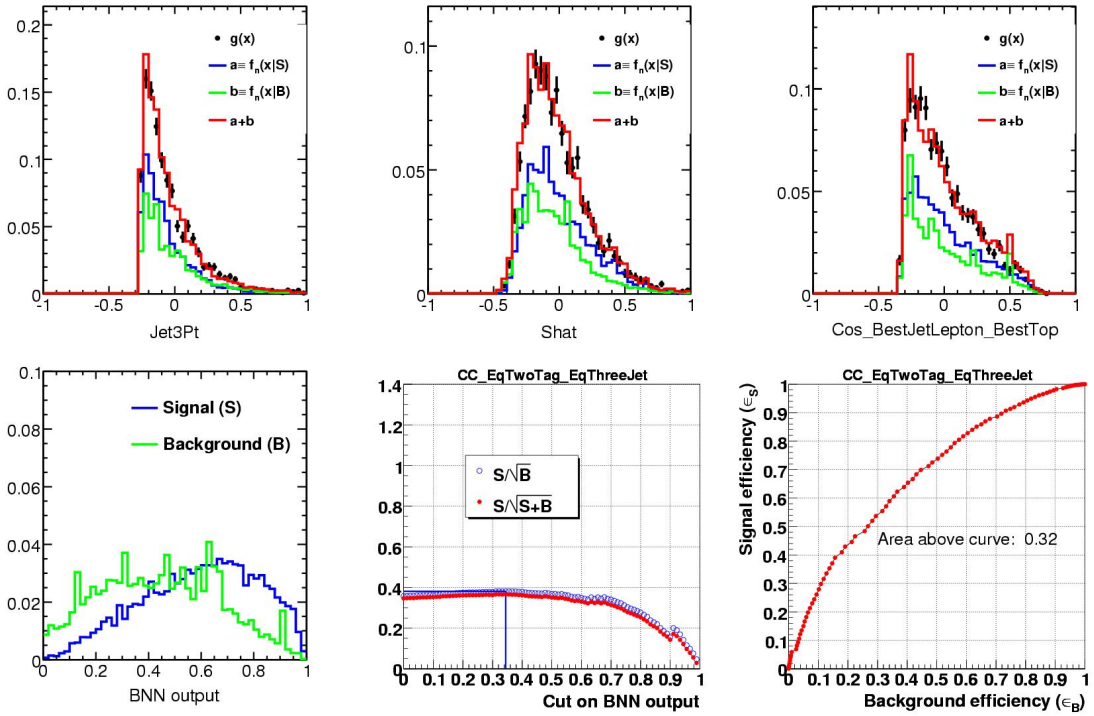


Figure 8.10: Rows: top = verification plots, bottom = BNN outputs normalized to unity (green: background, blue: signal), and the $S/\sqrt{B}$, $S/\sqrt{S+B}$, and signal efficiency (ϵ_S) versus background efficiency (ϵ_B) curves, for the electron, =2 tag, =3jet channel.

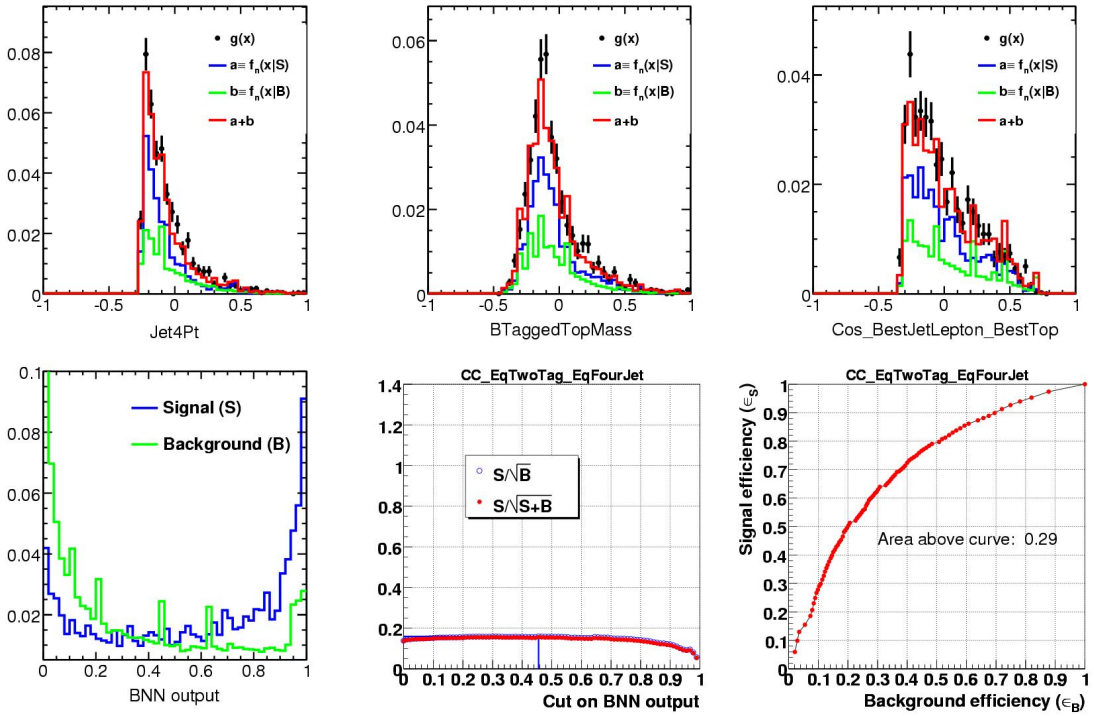


Figure 8.11: Rows: top = verification plots, bottom = BNN outputs normalized to unity (green: background, blue: signal), and the $S/\sqrt{B}$, $S/\sqrt{S+B}$, and signal efficiency (ϵ_S) versus background efficiency (ϵ_B) curves, for the electron, =2 tag, =4jet channel.

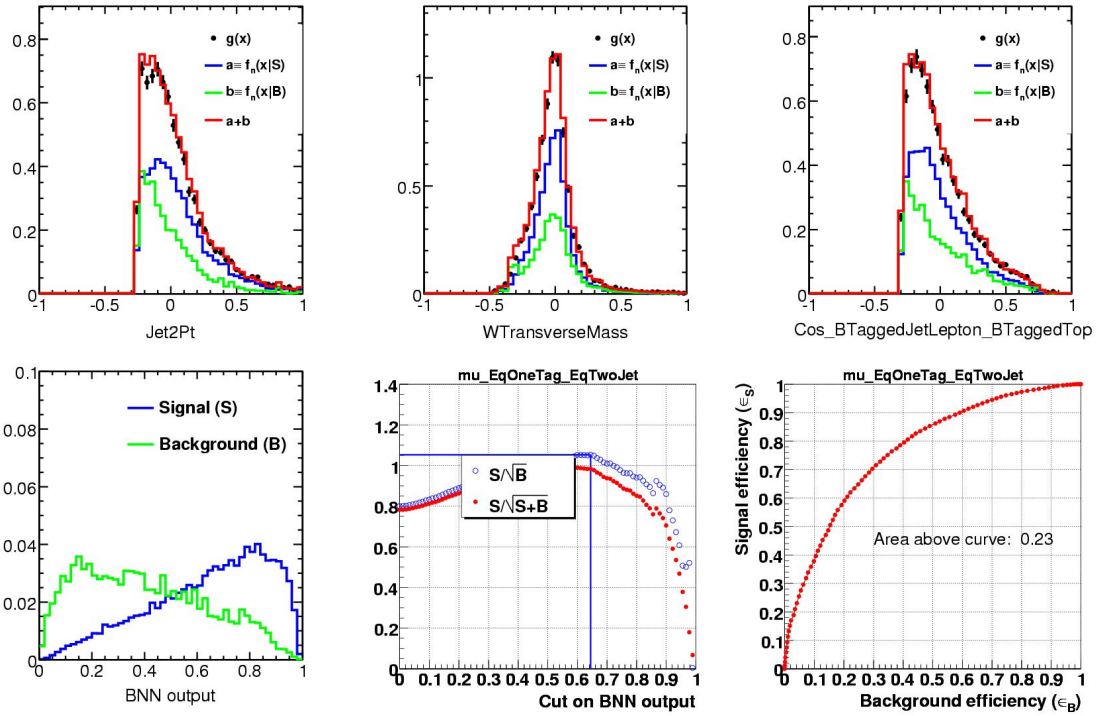


Figure 8.12: Rows: top = verification plots, bottom = BNN outputs normalized to unity (green: background, blue: signal), and the $S/\sqrt{B}$, $S/\sqrt{S+B}$, and signal efficiency (ϵ_S) versus background efficiency (ϵ_B) curves, for the muon, =1 tag, =2 jet channel.

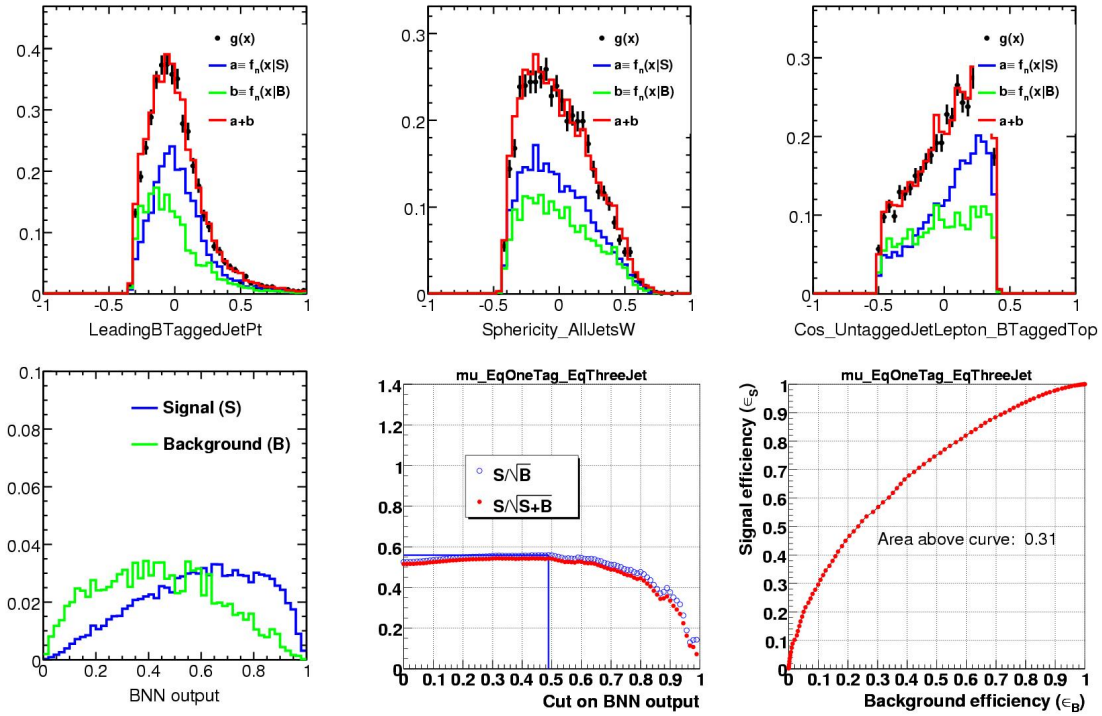


Figure 8.13: Rows: top = verification plots, bottom = BNN outputs normalized to unity (green: background, blue: signal), and the $S/\sqrt{B}$, $S/\sqrt{S+B}$, and signal efficiency (ϵ_S) versus background efficiency (ϵ_B) curves, for the muon, =1 tag, =3jet channel.

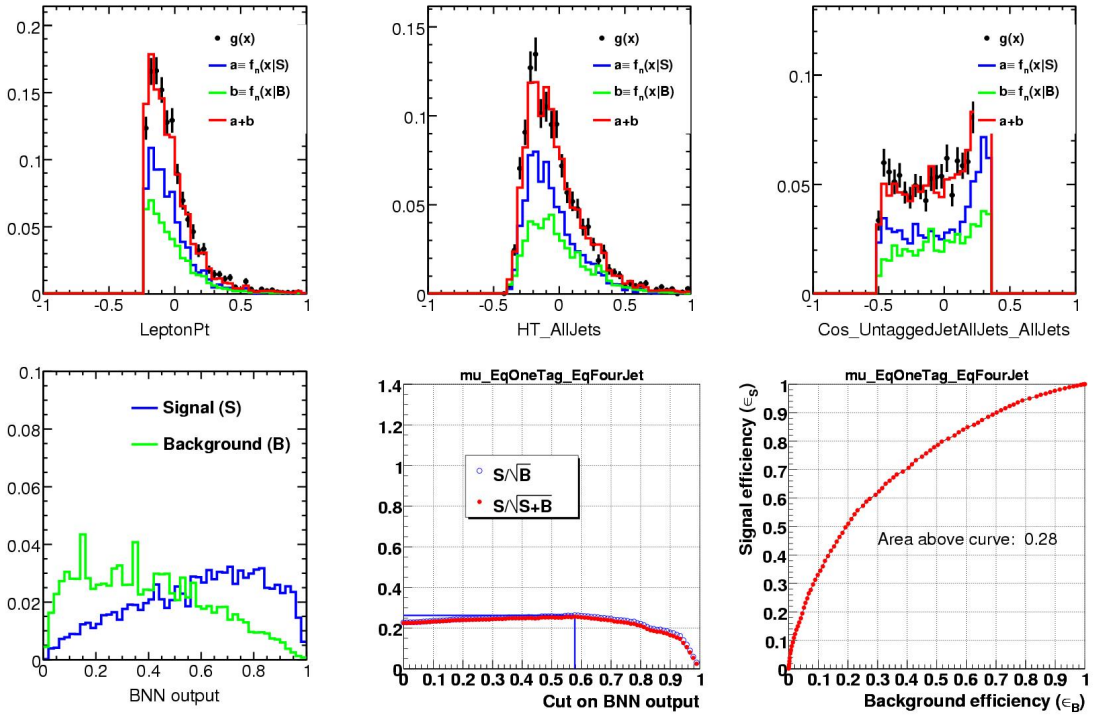


Figure 8.14: Rows: top = verification plots, bottom = BNN outputs normalized to unity (green: background, blue: signal), and the $S/\sqrt{B}$, $S/\sqrt{S+B}$, and signal efficiency (ϵ_S) versus background efficiency (ϵ_B) curves, for the muon, =1 tag, =4jet channel.

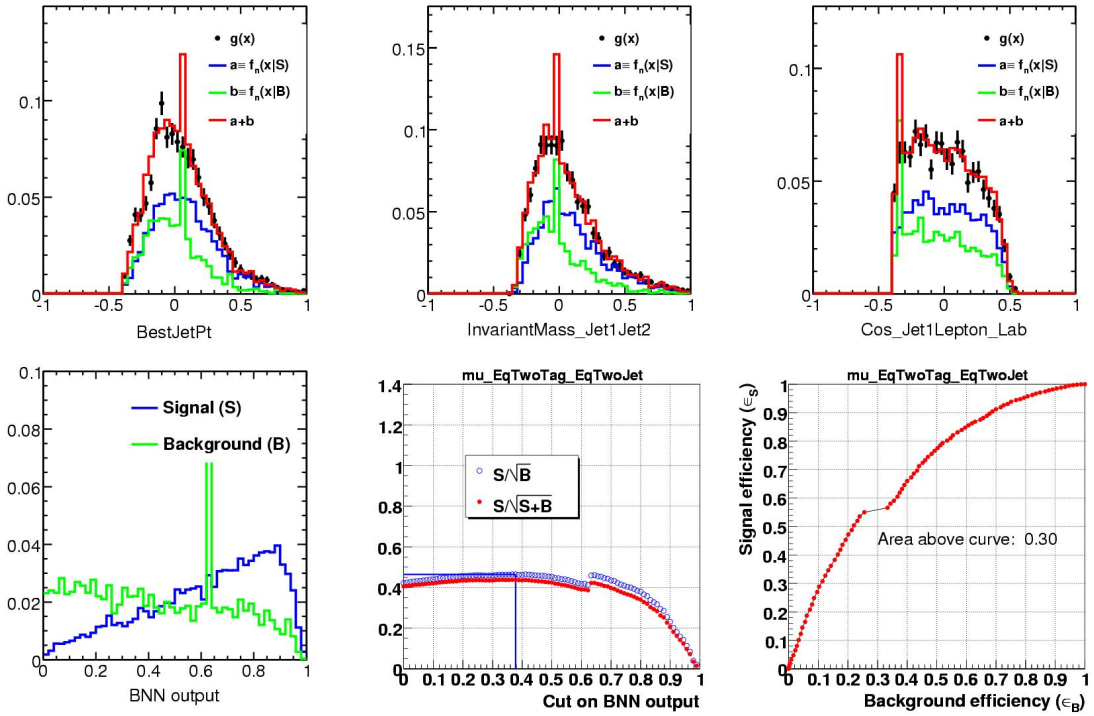


Figure 8.15: Rows: top = verification plots, bottom = BNN outputs normalized to unity (green: background, blue: signal), and the $S/\sqrt{B}$, $S/\sqrt{S+B}$, and signal efficiency (ϵ_S) versus background efficiency (ϵ_B) curves, for the muon, =2 tag, =2jet channel.

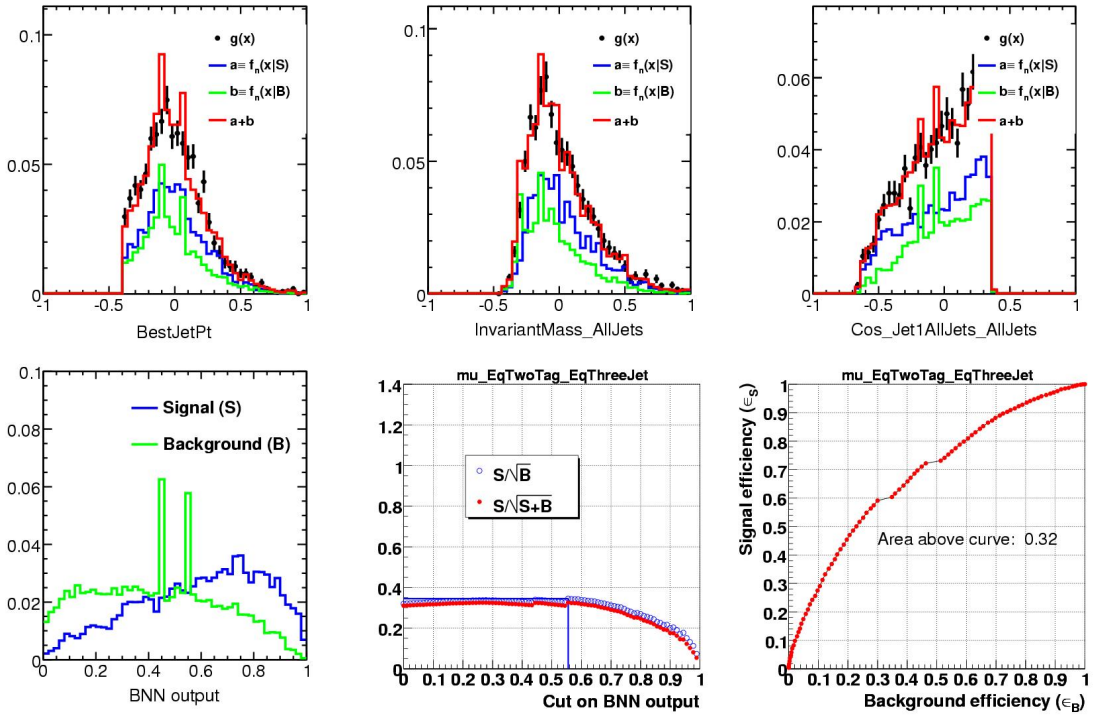


Figure 8.16: Rows: top = verification plots, bottom = BNN outputs normalized to unity (green: background, blue: signal), and the $S/\sqrt{B}$, $S/\sqrt{S+B}$, and signal efficiency (ϵ_S) versus background efficiency (ϵ_B) curves, for the muon, =2 tag, =3jet channel.

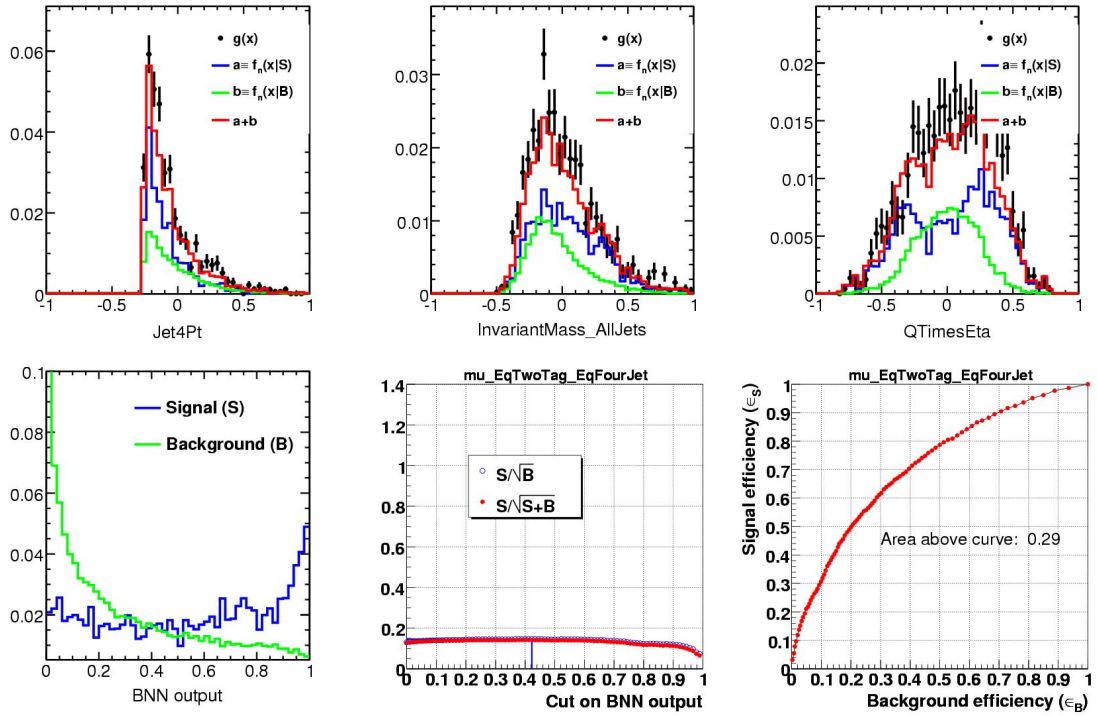


Figure 8.17: Rows: top = verification plots, bottom = BNN outputs normalized to unity (green: background, blue: signal), and the $S/\sqrt{B}$, $S/\sqrt{S+B}$, and signal efficiency (ϵ_S) versus background efficiency (ϵ_B) curves, for the muon, =2 tag, =4jet channel.

8.4 BNN Outputs

The BNN output distributions obtained by applying the BNN functions, $D(x)$, to the set of simulated events, and normalized to the observed integrated luminosity of 0.9 fb^{-1}, are shown in Figs. 8.18-8.19 for the twelve different analysis channels.

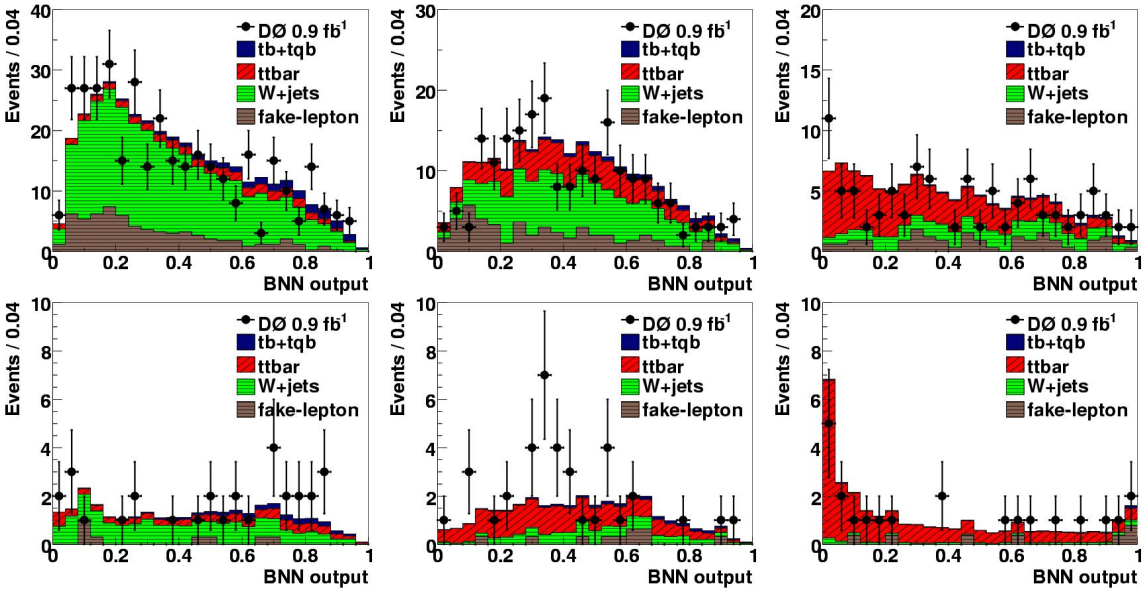


Figure 8.18: Bayesian neural network outputs for the electron channel. [Rows: top =1 tag, bottom =2 tags, columns: left =2 jets, middle =3 jets, right =4 jets.]

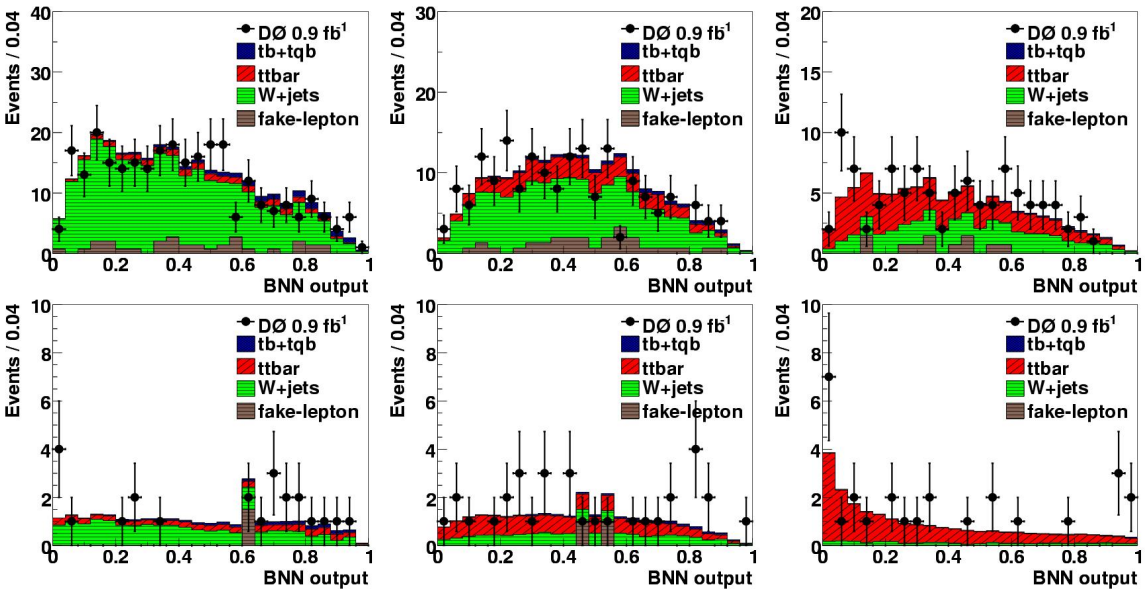


Figure 8.19: Bayesian neural network outputs for the muon channel. [Rows: top =1 tag, bottom =2 tags, columns: left =2 jets, middle =3 jets, right =4 jets.]

8.5 Cross Check Samples

In order to validate every step of the Bayesian neural network analysis, cross-check samples were used to decide whether the background model and data are in agreement after applying a Bayesian neural network function. Two cross-check samples are defined as follows: "W+jets" (2 jets, 1 b-tag, $H_T < 175$ GeV), and "$t\bar{t}$" (4 jets, 1 b-tag, $H_T > 300$ GeV). These samples are designed such that one of them has mostly W+jets events and almost no $t\bar{t}$, and the other is mostly $t\bar{t}$ and has almost no W+jets. This allows us to test whether or not each part of the background model is adequately described. Figures 8.20 and 8.21 show the Bayesian neural network output distributions for these cross-check samples. We see good agreement between the predicted background and the observed data in both the samples, for each of the electron and muon channels. We thus conclude that the background model describes the data well within uncertainties.

W+jets cross check samples

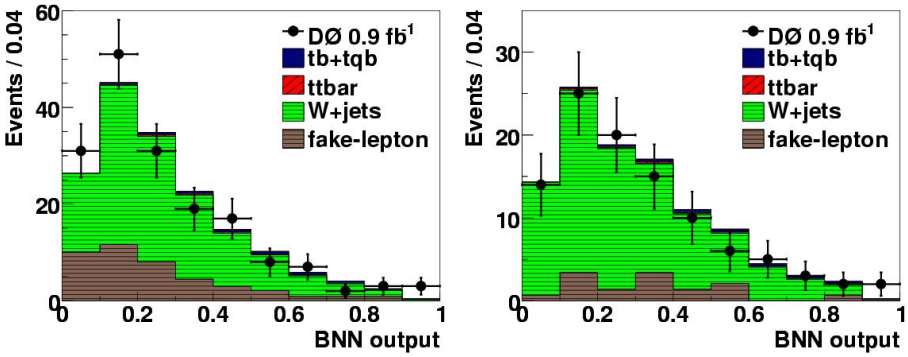


Figure 8.20: BNN outputs from W+jets cross check samples for the electron (left) and muon (right) channels.

$t\bar{t}$ cross check samples

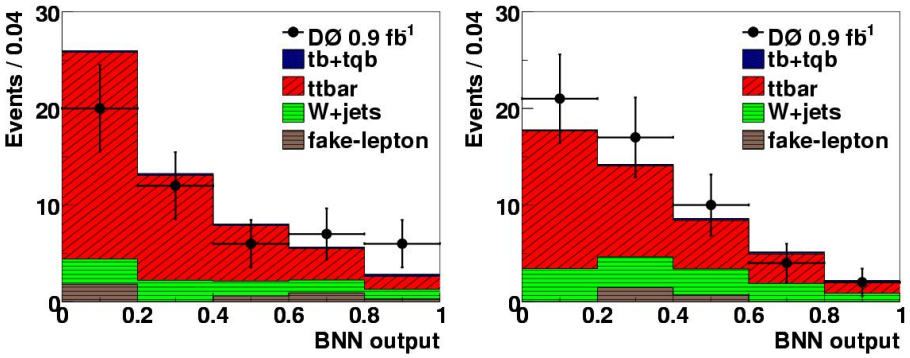


Figure 8.21: BNN outputs from $t\bar{t}$ cross check samples for the electron (left) and muon (right) channels.

8.6 Measuring Single Top Cross Section

We next apply Bayesian statistics as discussed in Ref. [84] to compute the posterior probability density of the single top cross section given an observed distribution of counts. We form a binned likelihood using the BNN output distributions and consider the sum of tb and tqb processes as the signal. All systematic effects described in are included taking into account the correlations. The measured $tb + tqb$ cross section is defined by the mode of the posterior density and a one-standard deviation uncertainty by the 68.3% area enclosed around the mode. An estimate of the sensitivity of our analysis is obtained from *expected* results for which we use a fake dataset, based on our signal and background model, in which the "observed" count in each bin of the BNN output distributions is set equal to the sum of the expected signal and background counts. When we use real data, that is, the observed distribution of counts, we call the corresponding results the *observed* results.

8.6.1 Optimization Studies

We studied different ways in which to use the BNN output distributions in order to get the best expected results. For the fake dataset, we generated pseudo-data events from a SM signal+background ensemble assuming a $tb + tqb$ cross section of 2.9 pb. When generating this ensemble, all systematic uncertainties were taken into account as discussed in Ref. [85]. We considered the following four ways to use the BNN output distributions from the different channels:

- Case A: 25 uniform bins across all twelve channels.
- Case B: 25 uniform bins across all channels, but discarding the four-jet channels. Since the S/B is the least in the four-jet channels, we wanted to check if these channels were contributing largely to noise rather than to a useful increase in acceptance.
- Case C: all twelve channels, with 25 uniform bins in the 1-tag channels and coarser (10) bins in the 2-tag channels, because these channels have very low yields.
- Case D: all twelve channels, with 25 uniform bins, but with a cut applied to the BNN output corresponding to maximum $S/\sqrt{B}$ for each channel, as obtained from the $S/\sqrt{B}$ curves in Figs. 8.6-8.17.

Overlay plots of the measured cross sections (σ_{tb+tqb}), uncertainties ($\Delta\sigma_{tb+tqb}$), and significances ($\sigma_{tb+tqb}/\Delta\sigma_{tb+tqb}$), including all systematic effects, for the above four cases are shown in Fig. 8.22. We see that there is negligible difference in distributions for the first three cases, whereas, there is a slight deterioration of performance in case D. We, therefore, considered case A, namely 25 uniform bins across all twelve channels for the rest of the analysis. We also checked that for this choice of binning, the truncation of the Gaussian distribution for the two shape-changing effects (JES and TRF) is negligible in the Bayesian analysis. The sampled Gaussian distributions are shown in Fig. 8.23.

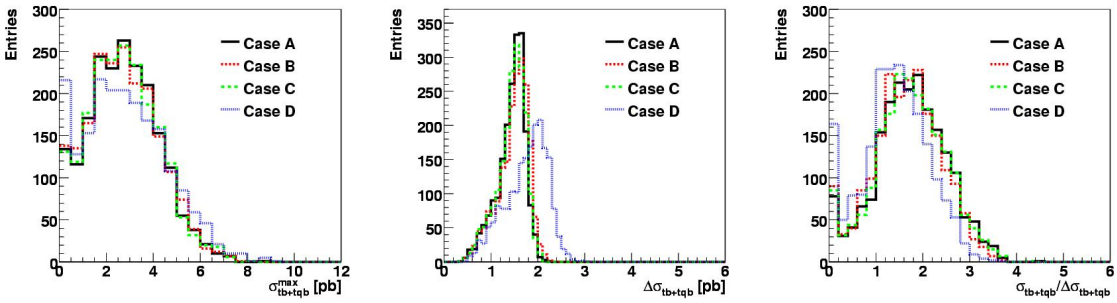


Figure 8.22: Distributions of the measured cross sections (σ_{tb+tqb}), uncertainties ($\Delta\sigma_{tb+tqb}$), and significances ($\sigma_{tb+tqb}/\Delta\sigma_{tb+tqb}$), including all systematic effects, for four different bin and channel combination options overlayed.

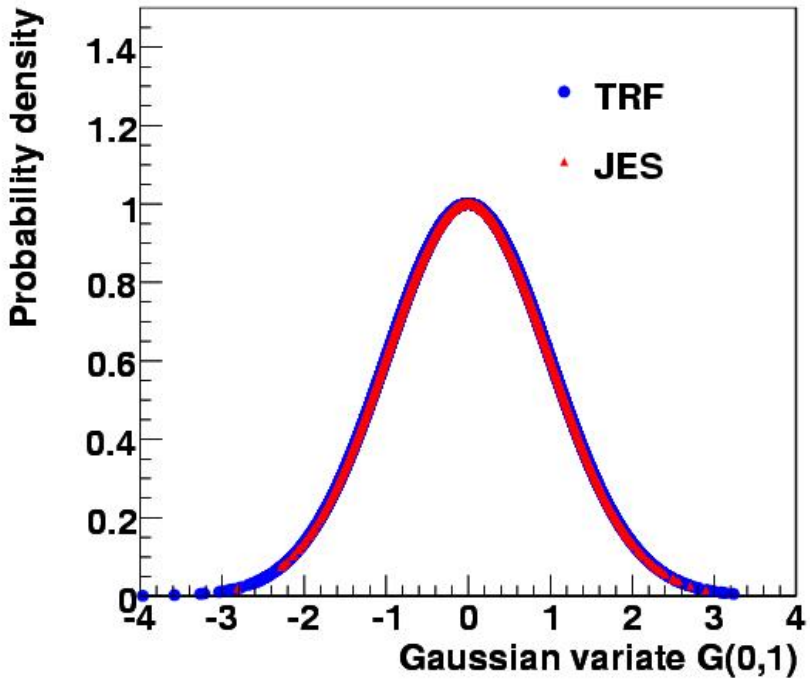


Figure 8.23: The Gaussian distributions (with zero mean and unit standard deviation) sampled for the JES and TRF systematics in the Bayesian analysis.

8.6.2 Determining the Bias

We used the best channel combination determined above, namely, all twelve channels with 25 uniform bins, to estimate the bias in our measurement of the single top cross section (σ_{tb+tqb}). For this, three ensembles of pseudo-data events were produced from the signal and background models, with single top cross sections in the range 0-10 pb, but assuming a SM ratio between the tb and tqb production cross sections ($0.88/1.98 = 0.44$). About 200 pseudo-data events were generated at the following $tb + tqb$ cross section values:

1. Ensemble A: $\sigma_{tb+tqb} = 7.9$ pb (2.76×SM)
2. Ensemble C: $\sigma_{tb+tqb} = 2.0$ pb (0.70×SM)
3. Ensemble D: $\sigma_{tb+tqb} = 5.7$ pb (2.00×SM)

When generating these sets of ensembles, the systematic effects were not included unlike what was done in the study described in section. But, when computing the posterior density for each pseudo-dataset, *all* systematic effects with correlations were included. Distributions of the measured cross sections for each ensemble are shown in Fig. 8.24.

The bias in the $tb + tqb$ cross section is determined from a straight-line fit to the mean of the distributions versus the input cross sections used to generate each ensemble type. We see from Fig 8.25 that the slope of the line is about unity with an offset close to zero. We thus conclude that the bias in our measurement is negligible and can be ignored henceforth.

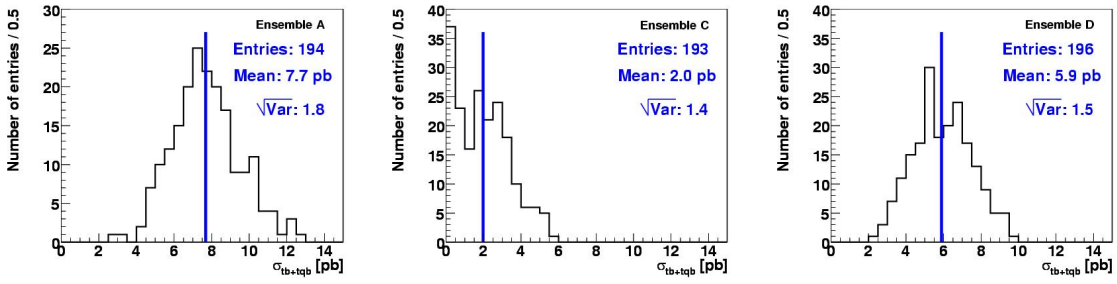


Figure 8.24: Distributions of the measured cross sections (peaks of Bayesian posterior density) from the non-SM ensembles. The arrow shows the mean of each distribution.

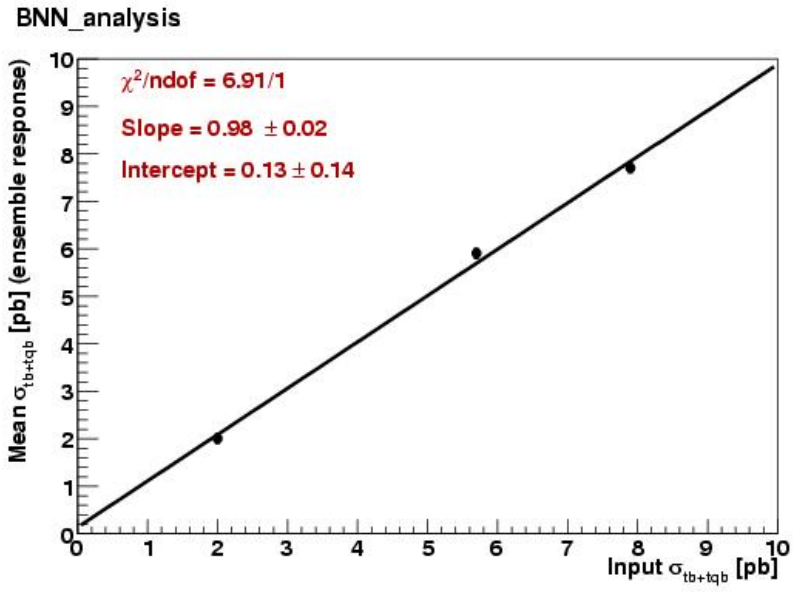


Figure 8.25: Ensemble response versus input cross section used to generate the ensemble. The ensemble response is obtained from the mean of the distributions in Fig. 8.24.

8.6.3 Expected Results

Having optimized the binning of the BNN output distributions using ensembles, and checked for biases in our measurements, we computed the expected results by setting the "observed" count in each bin of the BNN output distributions to the sum of expected SM signal+background. The results for the electron, muon and $e + \mu$ channels with all tag and jet multiplicities combined, are shown in Table 8.2 with no systematics and with all systematic effects included, with correlations. The resulting posterior density plots are shown in Fig. 8.26.

Table 8.2: Expected results for the electron, muon and combined channels.

All tag and jet multiplicities combined (using 25 uniform bins)		
Lepton channel	$\sigma \pm \Delta\sigma$	$\sigma/\Delta\sigma$
	No systematics	
Electron (e)	2.9 ± 1.5	2.0
Muon (μ)	2.9 ± 1.7	1.7
$e + \mu$	2.9 ± 1.2	2.5
	All systematics	
Electron (e)	2.7 ± 1.8	1.5
Muon (μ)	2.7 ± 2.2	1.2
$e + \mu$	2.7 ± 1.5	1.8

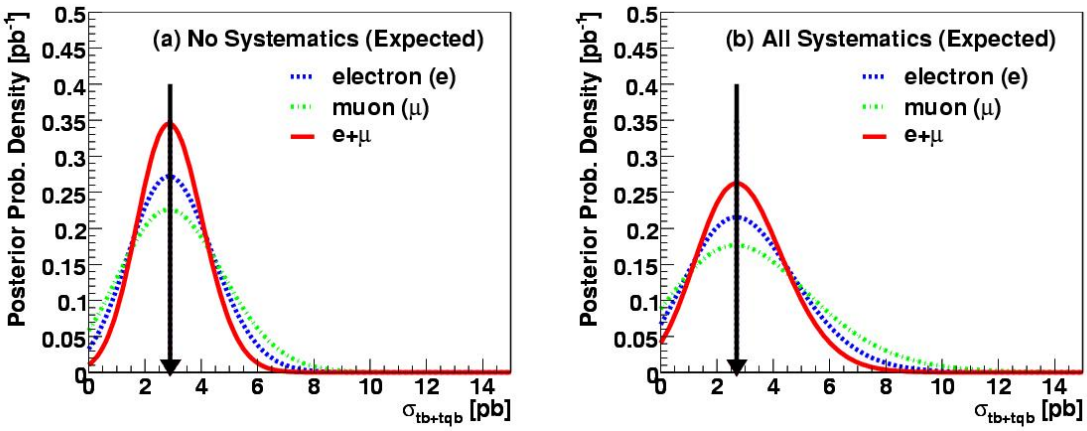


Figure 8.26: Distributions of the expected posterior probability density as a function of the $tb + tqb$ single top cross section with (a) no systematics, and (b) all systematics, and combining the tag and jet multiplicities for different lepton combinations.

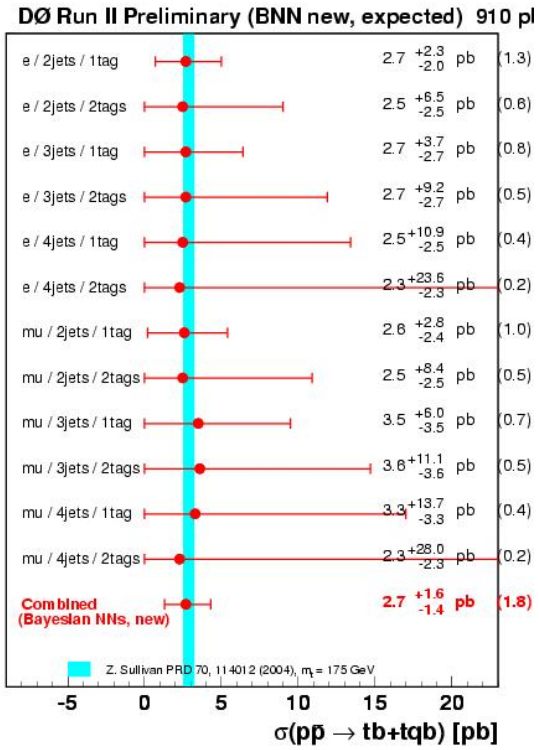


Figure 8.27: Results of the expected measurements in the twelve channels individually and combined. The numbers in parentheses in the right-most column are the expected significances defined as $\sigma_{tb+tqb}/\Delta\sigma_{tb+tqb}$.

8.6.4 Observed Results

Finally, we present the single top cross section measurements using real data. The results for the electron, muon and $e+\mu$ channels with all tag and jet multiplicities combined, are shown Table 8.3 with no systematics and including all systematic effects. The resulting posterior density plots are shown in Fig. 8.28. A comparison of the observed measurements across the twelve channels individually, and combined, is shown in Fig. 8.29.

Table 8.3: Observed results for the electron, muon and combined channels.

All tag and jet multiplicities combined (using measured data, and 25 uniform bins)		
Lepton channel	$\sigma \pm \Delta\sigma$	$\sigma/\Delta\sigma$
	No systematics	
Electron (e)	3.1 ± 1.4	2.2
Muon (μ)	3.4 ± 1.7	2.0
$e + \mu$	3.2 ± 1.1	2.9
	All systematics	
Electron (e)	4.6 ± 2.0	2.3
Muon (μ)	4.5 ± 2.3	2.0
$e + \mu$	4.4 ± 1.5	3.0

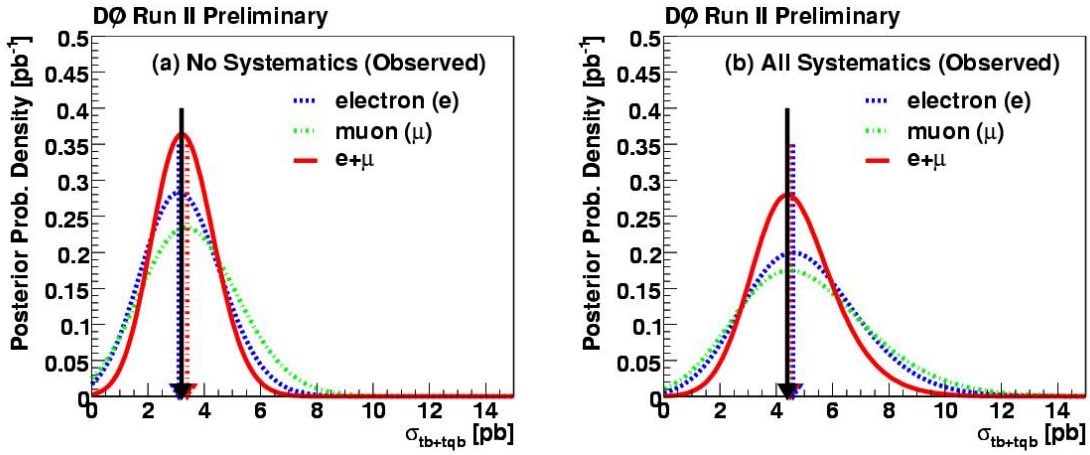


Figure 8.28: Distributions of the posterior probability density as a function of the $tb + tqb$ single top cross section with (a) no systematics, and (b) all systematics, using the measured data and combining the tag and jet multiplicities for different lepton combinations.

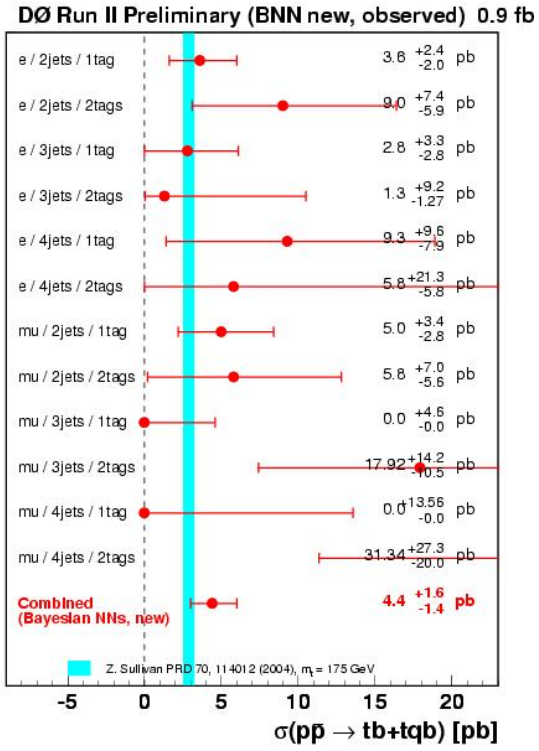


Figure 8.29: Measurements in the twelve channels individually, and combined.

8.7 Probabilities and Signal Significance

Having measured the expected and observed single top ($tb + tqb$) cross sections, we use an ensemble of background-only samples, generated with all systematic effects included, to compute p-values. We define the p-value to be the probability to obtain a measured cross section equal to or higher than a reference value. Figure 8.30 shows the distribution of cross sections from the background-only ensemble. Using the SM signal cross section of 2.9pb, we obtain a p-value of 1.6%, which corresponds to a 2.2 standard deviation expected significance. Using the observed measurement of 4.4pb, we obtain a p-value of 0.08%, which corresponds to a significance of 3.2 standard deviations.

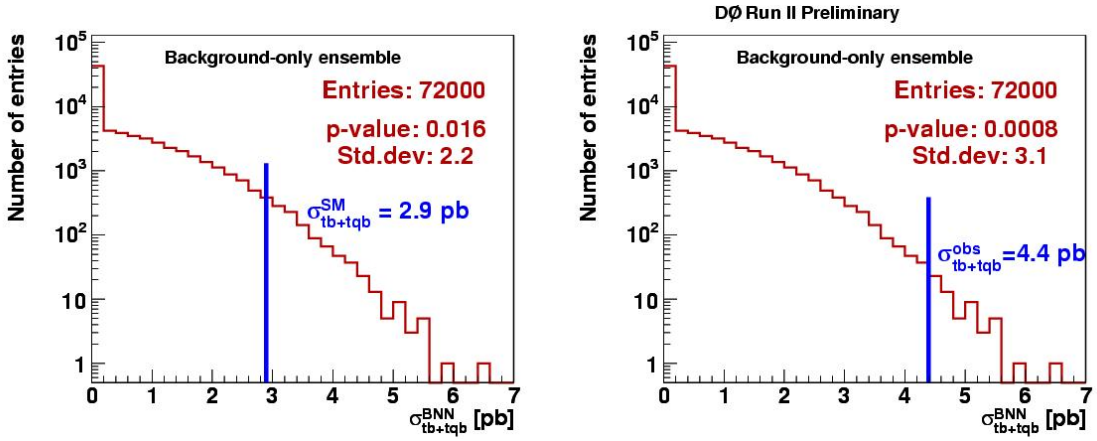


Figure 8.30: Distribution of cross sections from background-only ensemble with full systematics included.

We also use the SM signal ensemble to compute the probablity of measuring the single top cross section equal to 4.4pb or higher, which is a rough measure of how likely it is to measure a cross section of 4.4pb if the single top cross section were truly 2.9 pb. It is a p-value associated with the (new) null hypothesis: the single top cross section is 2.9 pb. Figure 8.31 shows the distribution of cross sections from the ensemble, from which we obtain the above probability as 13.7%, which corresponds to a 1.1 standard deviation fluctuation. We therefore have no compelling reason to reject the hypothesis that the single top cross section is as predicted by the Standard Model.

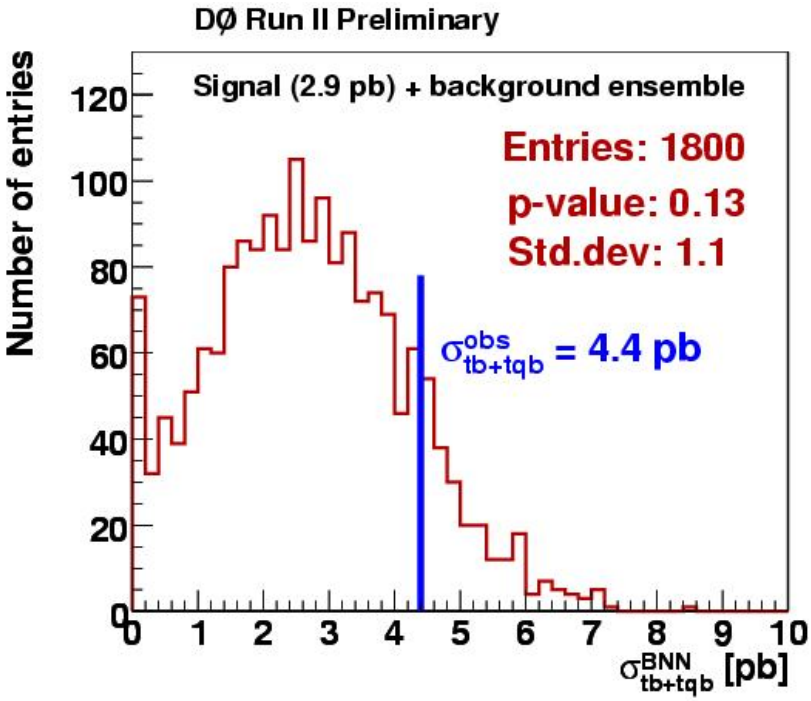


Figure 8.31: Distribution of cross sections from SM signal+background ensemble with full systematics included.

8.8 Summary and Conclusions

We used Bayesian neural networks to separate single top quark signals from background in a sample of lepton+jets events selected from nearly 1 fb^{-1} of Run II data. The BNN output distributions across twelve independent channels were combined using a binned likelihood and the single top ($tb + tqb$) cross section measured using a Bayesian method. We obtained the following result:

$$\sigma\left(p\bar{p} \rightarrow tb + tqb + X\right) = 4.4^{+1.6}_{-1.4} \text{ pb}.$$

This results in a p-value of 0.08%, corresponding to a 3.2 standard deviation significance. The p-value from the SM value of 2.9 pb for the signal cross section is 1.6%, which corresponds to a 2.2 standard deviation expected significance. Figure 8.32 shows the observed BNN distribution summed over all 12 channels, superimposed on the summed signal + background model. We note the good agreement between observed data and model.

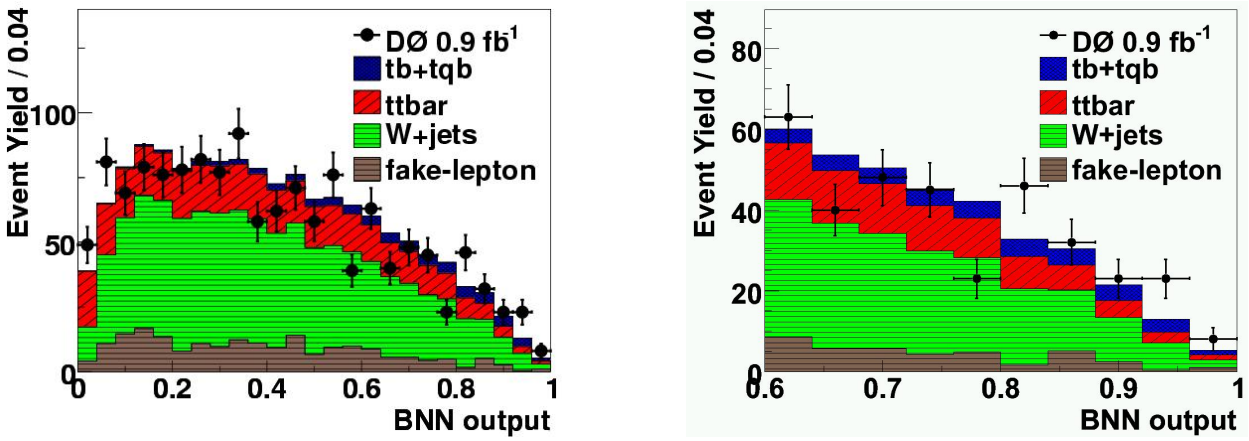


Figure 8.32: The observed BNN output distribution summed over all 12 channels superimposed on the summed signal + background model. The plot on the right is a zoom of the region near a BNN output of 1.

APPENDIX A

Discrepancy Measures

The standard way to verify that the modeling of a variable is satisfactory is to compare its distribution, as predicted by the model, with the observed one. We assume, as usual, that these distributions are given as histograms and that the hypothesis to be rejected is that the two histograms agree. One defines a measure of discrepancy d between the two histograms such that larger values of d signify greater evidence against the hypothesis. For an observed discrepancy $d_{\rm obs}$, one computes the p-value = Prob($d > d_{\rm obs}$|Model). If the p-value is judged to be too small, one rejects the hypothesis that the two histograms agree and therefore that the variable is well-modeled. This calculation of course requires knowledge of the distribution of the discrepancy measure d under the assumed model.

The two most commonly used discrepancy measures for 1-dimensional histograms are the Kolmogorov-Smirnov (K-S) and χ^2 statistics. The K-S statistic is defined by $d = \max|c(d|\text{Data}) - c(d|\text{Model})|$, where $c(d|\text{Data})$ and $c(d|\text{Model})$, respectively, are the observed and predicted cumulative distribution functions (cdf) of the discrepancy measure d. The χ^2 statistic is given by $\chi^2 = \sum_i (N_i - f_i)^2/f_i$ where N_i is the observed count in the $i^{\rm th}$ bin and f_i is the expected count. There are standard routines to compute p-values for both statistics. However, the p-value reported for the K-S discrepancy measure is reliable only if the measure is calculated using un-binned data. For the χ^2 p-value to be valid the counts per bin must have an approximately Gaussian distribution.

We can avoid both restrictions, however, if we use the *exact* distributions of the discrepancy measures. The distributions can be derived empirically using an ensemble of $K(= 20,000)$ histograms sampled from the model histogram with each bin count sampled independently according to $N_i \sim \text{Poisson}(N_i|f_i)$. The corresponding ensemble of K discrepancies are then sorted into ascending order. Given an observed discrepancy $d_{\rm obs}$,

the p-value is given by $(K-n)/K$, where n is the ordinal value of the largest value of d, in the ordered sequence, for which $d \leq d_{\rm obs}$ [1]. Figure A.1 shows the K-S p-values, as calculated by us, for each variable pertaining to the (electron, 1-tag, 2-jet)-channel. Similar plots are available for the other eleven channels.

Local Discrepancies

The K-S and χ^2 discrepancy measures are only two of (presumably) infinitely many measures one could calculate, each quantifying a different kind of discrepancy between two histograms. Since our goal, ultimately, is to find an excess over background, we were motivated to explore a different measure of discrepancy. We define a local discrepancy as a contiguous sequence of bins in which the observed count in each bin differs, *in the same direction*, from the expected count. Any such deviation is potentially evidence against the hypothesis that the model and data agree. However, since we expect many local discrepancies to arise purely by chance, even if the hypothesis were true, we need a way to assess whether any one of them provides significant evidence against it.

Let C be the observed count in a local discrepancy and F the corresponding expected count. In the spirit of the K-S measure, we take $d = |C-F|$ as the measure of the discrepancy between the model and the observed histogram. We do not distinguish between an excess or a deficit, hence the absolute value in the definition of the discrepancy measure d.

The most obvious way to pick the most significant local discrepancy is to pick the one with the largest value of the discrepancy measure d, thus defining the discrepancy measure $D_{\rm max} = \max|C-F|$. A possible shortcoming of this measure is that a large value of $d = D_{\rm max}$ may prove to be less significant in a probabilistic sense than a smaller value. It is perhaps more reasonable to pick the local discrepancy that is the least likely under the hypothesis being tested (generally, referred to as the null). We take least likely to mean that the p-value associated with the local discrepancy and defined by p-value $\equiv \Pr(x \geq C|F) = \sum_{x \geq C} \text{Poisson}(x|F)$ is the smallest of all local discrepancies. Note that if we have a deficit, that is, $C < F$, we define the p-value by p-value $\equiv \Pr(x \leq C|F) = \sum_{x \leq C} \text{Poisson}(x|F)$. This defines the discrepancy measure $D_{\rm minp} = |C-F|(\min p\text{-value})$.

Consider the table below, which shows details of two local discrepancies from a compar-

[1] We found it useful to compute the ordinal value n using a binary search. This was done to reduce the computation time of the p-values.

ison of two histograms.

F	C	$d = \lvert C - F \rvert$	p-value
12.91	26	13.09	0.0009
46.92	29	17.92	0.0034

The second local discrepancy has the larger d, but since its p-value is greater than that of the first it represents a deviation from the model histogram that is more likely under the hypothesis that the two histograms agree and hence less significant. Therefore, although the first discrepancy is smaller we consider it to be the more significant discrepancy.

The distribution of the most significant local discrepancy and associated p-values are estimated using the same algorithm as for the K-S and χ^2 discrepancy measures. Some results are shown in Fig. 8.1.

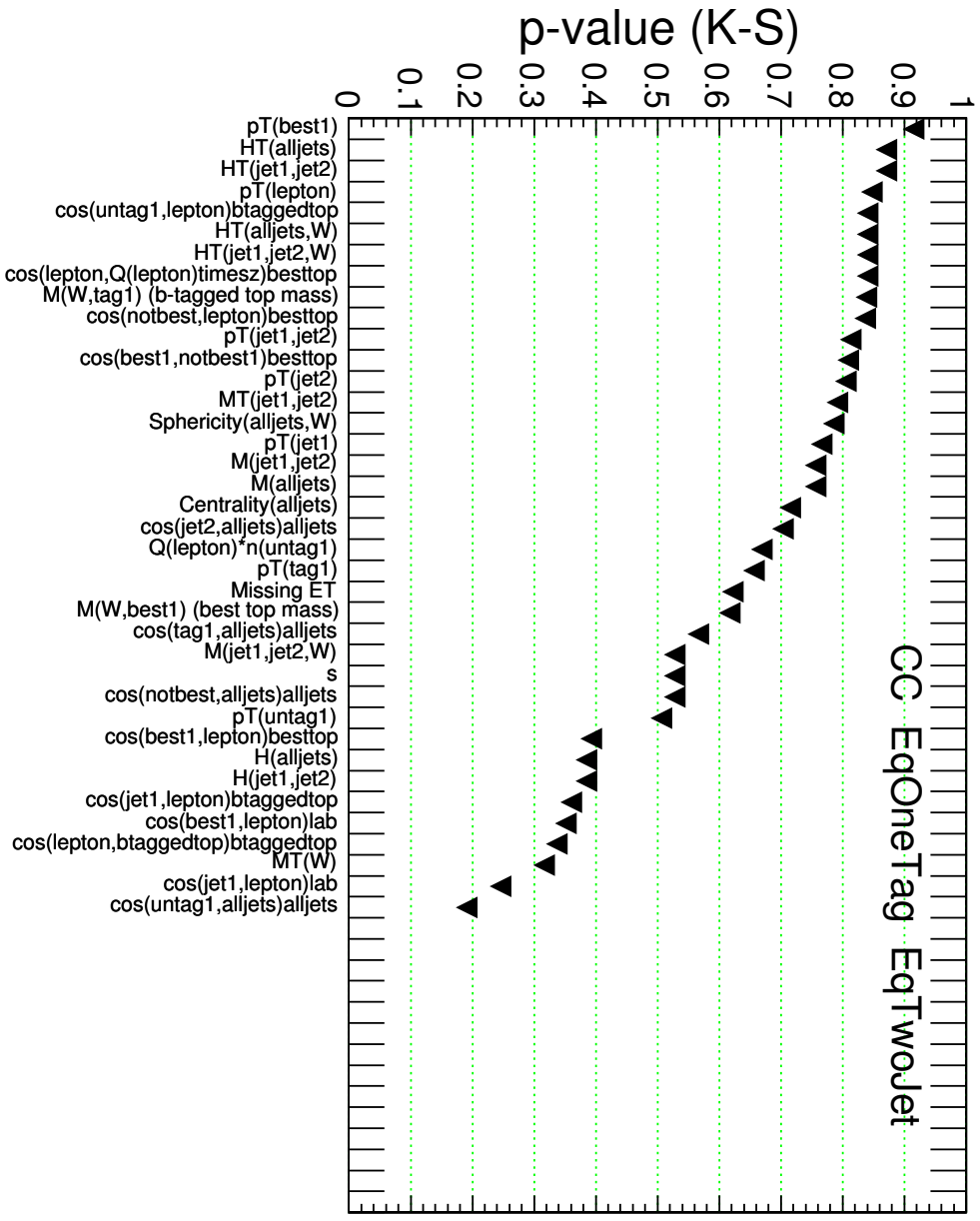


Figure A.1: (Electron,1-tag,2-jet)-channel variables ordered according to quality of modeling, quantified by the K-S p-value. The larger the p-value, the less reason one has to reject the hypothesis that the associated variable is poorly modeled, provided of course the kind of discrepancy measured by the K-S statistic is the most appropriate. A different discrepancy measure could yield a different conclusion.

APPENDIX B

Prior Width Estimates Using JETNET

For each of the 12 channels, the standard deviations of the (Gaussian) priors for the BNN were held fixed during the MCMC sampling of the posterior density, $p(w|T)$, of the network parameters. This was done to keep the parameter values from growing too large. In the previous BNN analysis, these parameters became very large because the priors were allowed to adapt to the noise level of the training data. This is a reasonable thing to do, in principle. However, if the training data are particularly noisy an adaptive prior can cause problems, as we discovered. By keeping the standard deviations of the priors fixed, at some appropriate values, this problematic behavior was avoided.

Our goal is to produce smooth, accurate, approximations to the discriminant $D(x)$ in Eq. (8.2). Small values of the network parameters tend to produce smoother approximations. If the values are too large this yields jagged and therefore poor approximations. However, if the weights are too small one risks losing discrimination power. Therefore, some reasonable choice for the characteristic scale of the network parameter values must be made. The appropriate scales were derived from a *single* network trained, using the back-propagation algorithm (a variant of gradient descent) as implemented in the JETNET program, for each of the twelve channels separately. The network structure of the JETNET-trained networks was identical to that used for the BNNs, as were the training data. Their output distributions are shown in Figs. B.1 and B.2.

For each of the 12 JETNET-trained networks (one per channel) the RMS with respect to zero for each of the three groups of network parameters—input weights, output weights, and hidden and output biases—was calculated. For each channel, the three RMS values served as the standard deviations of each of the three corresponding sets of zero mean Gaussians that served as priors for the BNN. The three standard deviations used, for each of the twelve

channels, are shown in Tables B.1 and B.2.

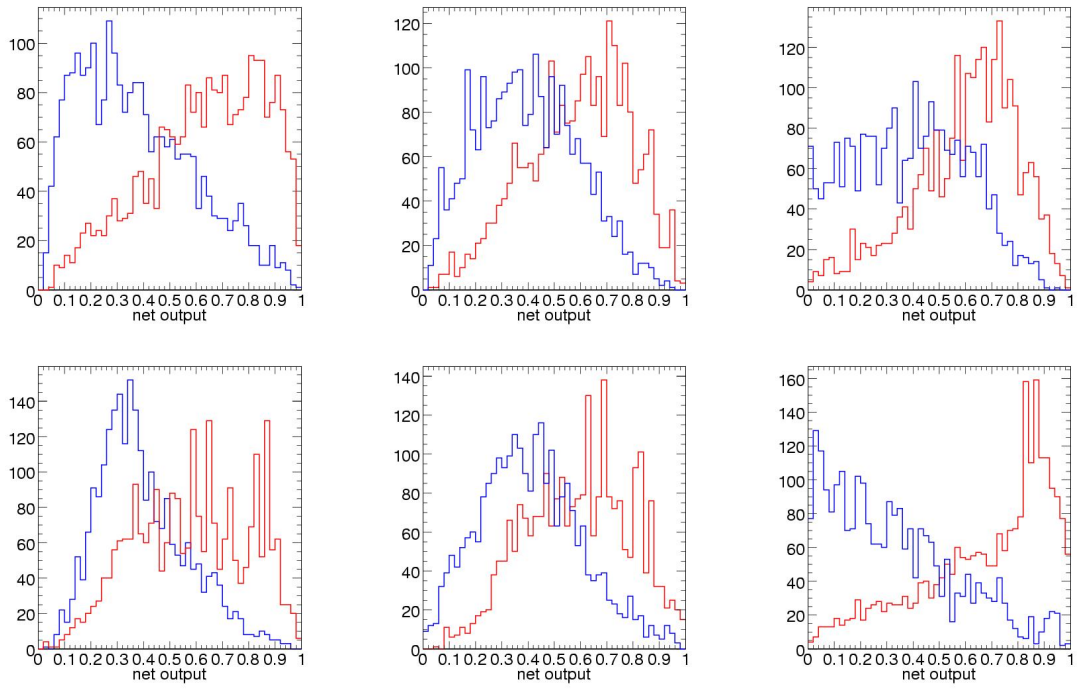


Figure B.1: Neural network output distributions for the electron channel as obtained using the JETNET neural network program. The blue histogram is for the sum of all backgrounds, and the red one is for the $tb + tqb$ signal. Rows: top = EqOneTag, bottom = EqTwoTag; columns: left = EqTwoJet, middle = EqThreeJet, right = EqFourJet.

Table B.1: The standard deviations for the *input-to-hidden* weights, σ_u, *hidden-to-output* weights, σ_v, and all biases, $\sigma_{a,b}$ obtained from a single neural network, JETNET, for the electron channel.

	Electron channel					
	EqOneTag			EqTwoTag		
	EqTwoJet	EqThreeJet	EqFourJet	EqTwoJet	EqThreeJet	EqFourJet
σ_u	0.24	0.22	0.21	0.24	0.20	0.26
σ_v	0.80	0.75	0.71	0.75	0.62	0.98
$\sigma_{a,b}$	0.41	0.38	0.26	0.37	0.24	0.30

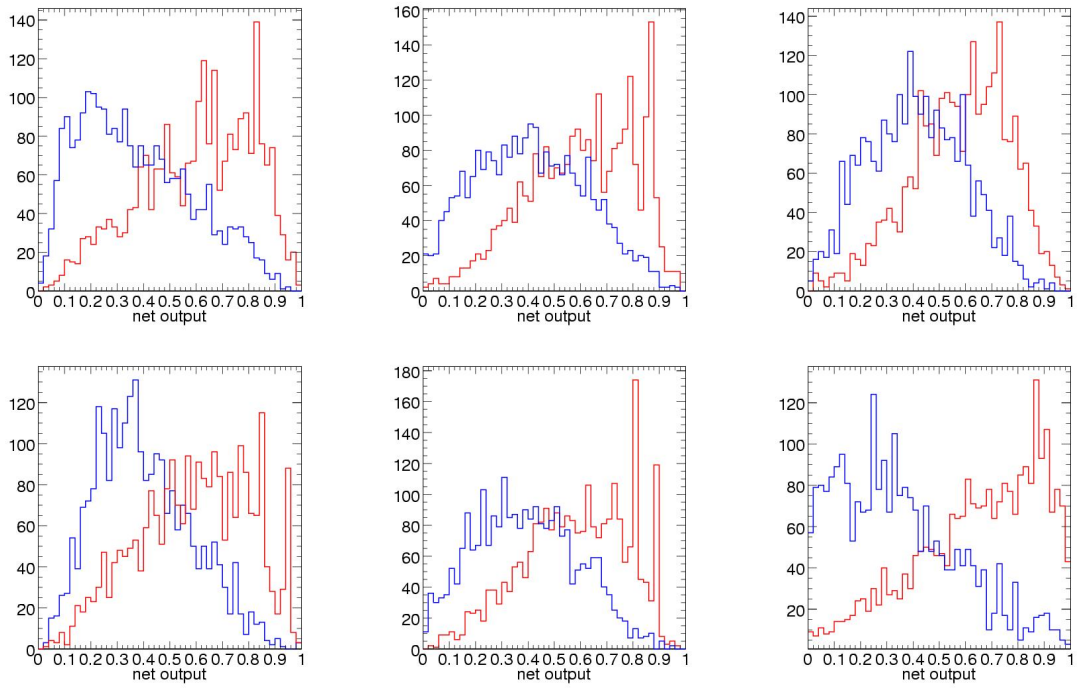


Figure B.2: Neural network output distributions for the muon channel as obtained using the single network JETNET algorithm. The blue histogram is for the sum of all backgrounds, and the red one is for the $tb + tqb$ signal. Rows: top = EqOneTag, bottom = EqTwoTag; columns: left = EqTwoJet, middle = EqThreeJet, right = EqFourJet.

Table B.2: The standard deviations for the *input-to-hidden* weights, σ_u, *hidden-to-output* weights, σ_v, and all biases, $\sigma_{a,b}$ obtained from a single neural network, JETNET, for the muon channel.

	Muon channel					
	EqOneTag			EqTwoTag		
	EqTwoJet	EqThreeJet	EqFourJet	EqTwoJet	EqThreeJet	EqFourJet
σ_u	0.22	0.22	0.22	0.25	0.22	0.29
σ_v	0.77	0.73	0.75	0.80	0.75	0.87
$\sigma_{a,b}$	0.35	0.31	0.26	0.30	0.44	0.36

APPENDIX C

Plots After Selection

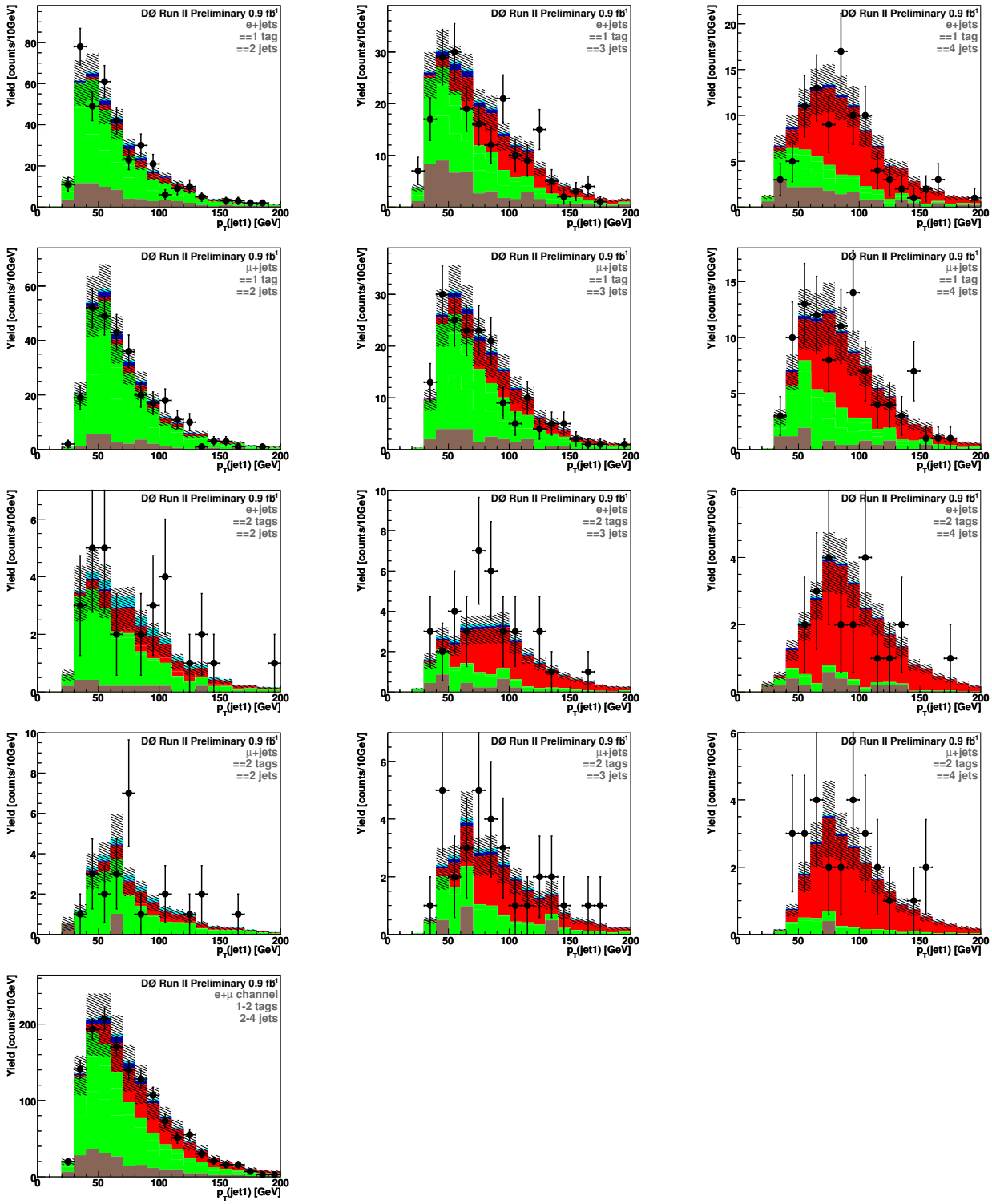


Figure C.1: The transverse momentum of jet 1 for events with two jets (left column), three jets (center column), and three jets (right column), for electron/1tag (first row), muon/1tag (second row), electron/2tags (third row), and muon/2tags (fourth row). The plot at the bottom of the figure shows the distribution for the electron and muon channels combined, for 2, 3, and 4 jets combined, and for 1 and 2 tags combined.

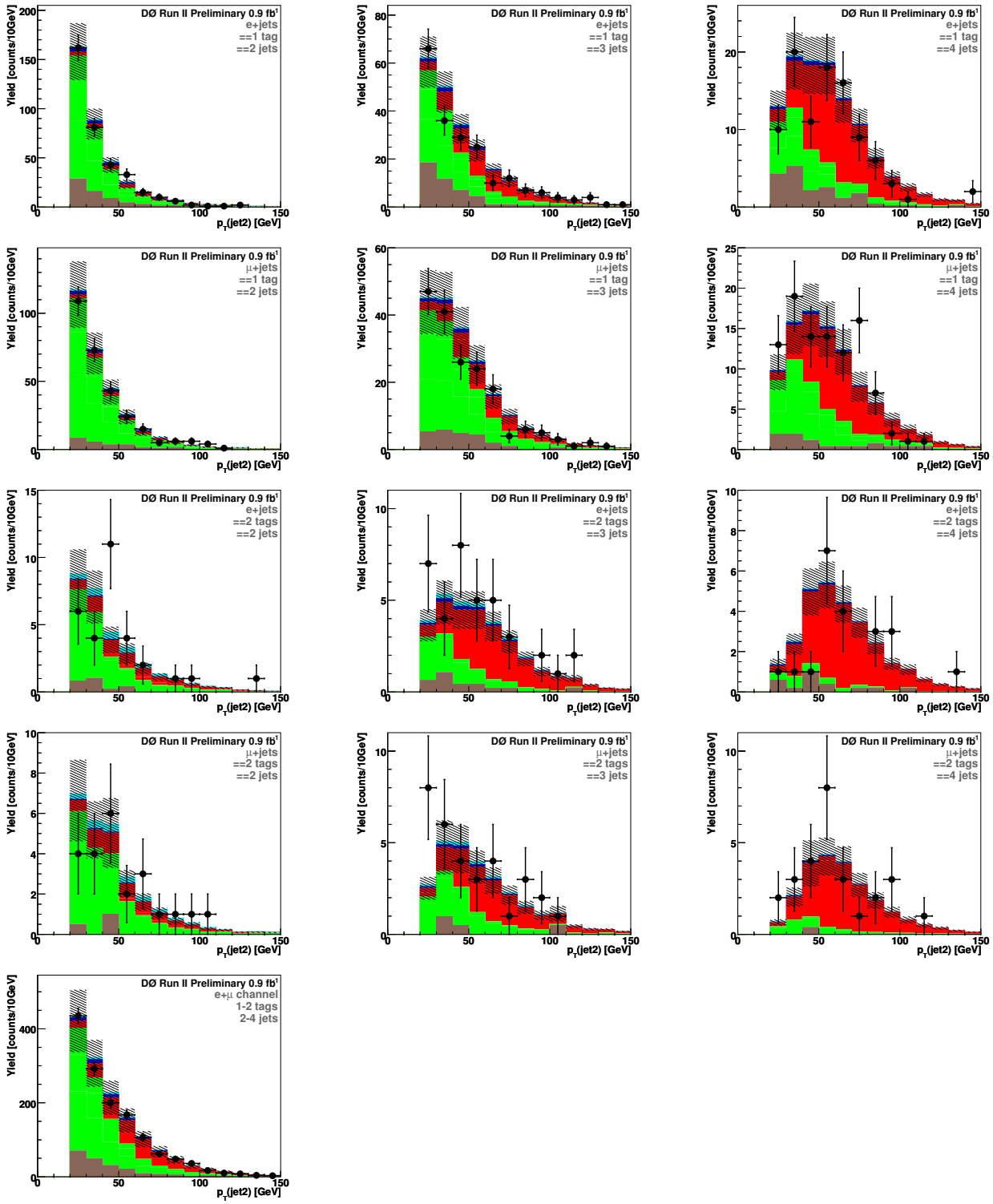


Figure C.2: The transverse momentum of jet 2 for events with two jets (left column), three jets (center column), and three jets (right column), for electron/1tag (first row), muon/1tag (second row), electron/2tags (third row), and muon/2tags (fourth row). The plot at the bottom of the figure shows the distribution for the electron and muon channels combined, for 2, 3, and 4 jets combined, and for 1 and 2 tags combined, i.e., all the channels we have used for this analysis.

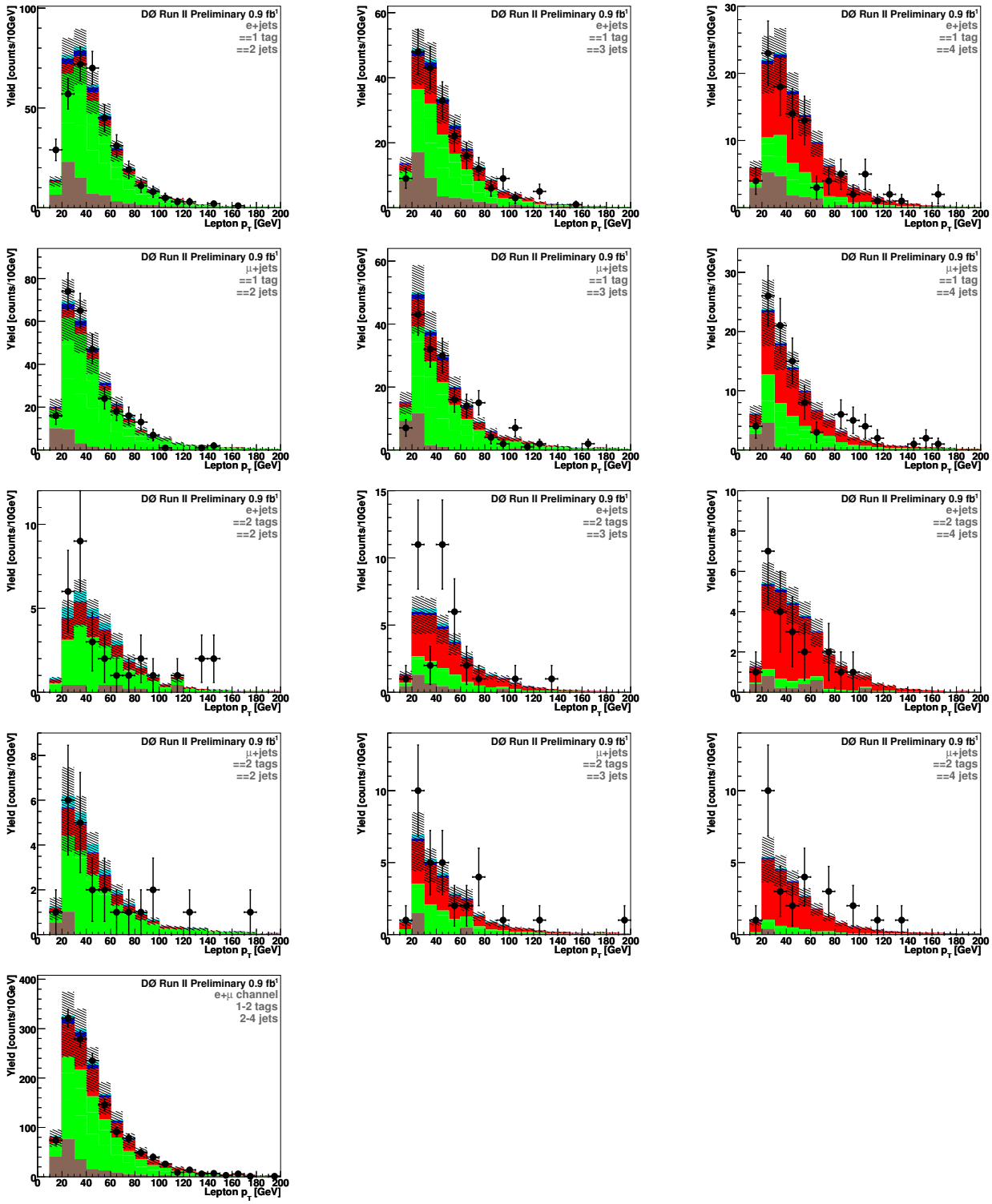


Figure C.3: The transverse momentum of the electron or muon for events with two jets (left column), three jets (center column), and three jets (right column), for electron/1tag (first row), muon/1tag (second row), electron/2tags (third row), and muon/2tags (fourth row). The plot at the bottom of the figure shows the distribution for the electron and muon channels combined, for 2, 3, and 4 jets combined, and for 1 and 2 tags combined, i.e., all the channels we have used for this analysis.

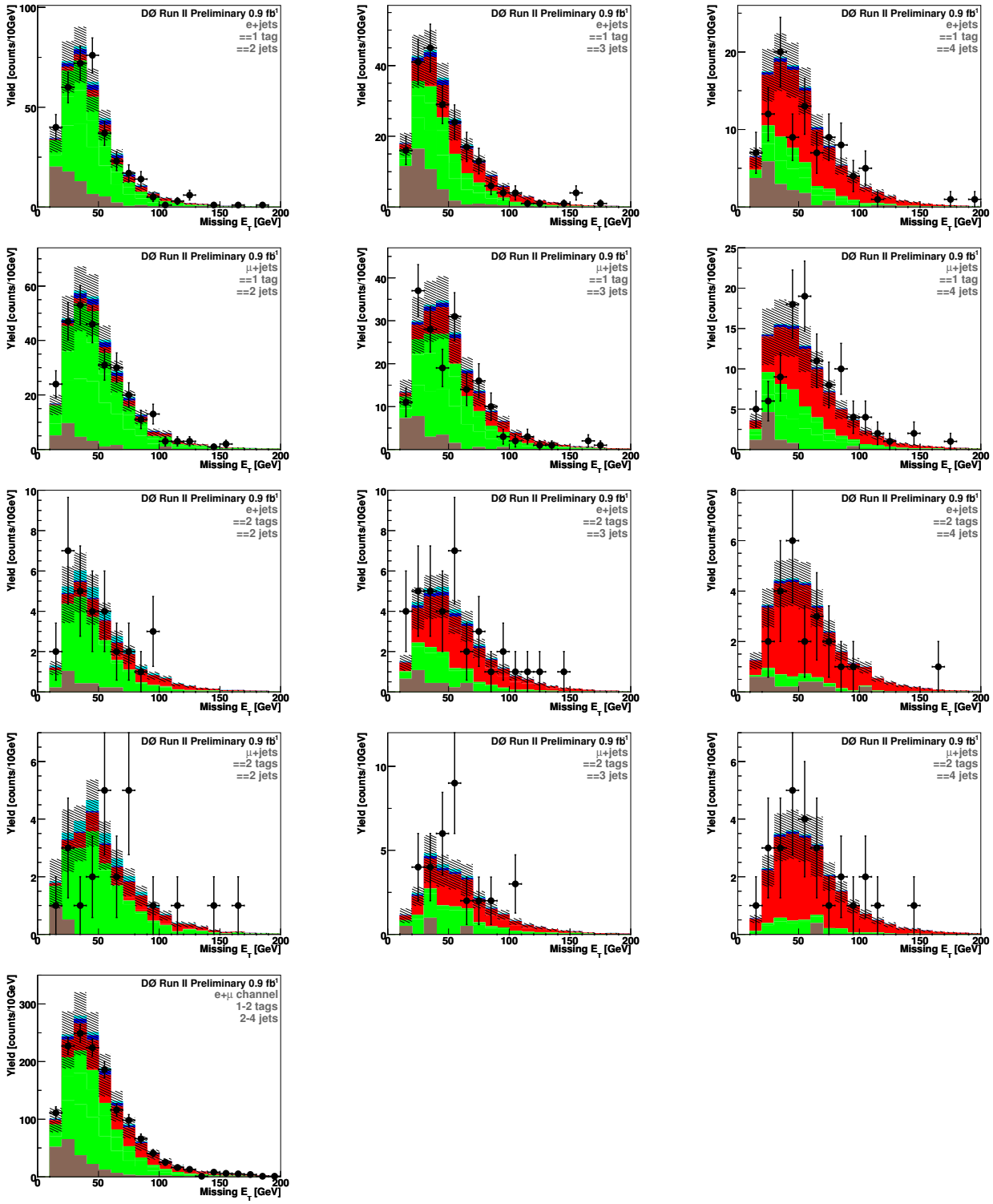


Figure C.4: The missing transverse energy for events with two jets (left column), three jets (center column), and three jets (right column), for electron/1tag (first row), muon/1tag (second row), electron/2tags (third row), and muon/2tags (fourth row). The plot at the bottom of the figure shows the distribution for the electron and muon channels combined, for 2, 3, and 4 jets combined, and for 1 and 2 tags combined, i.e., all the channels we have used for this analysis.

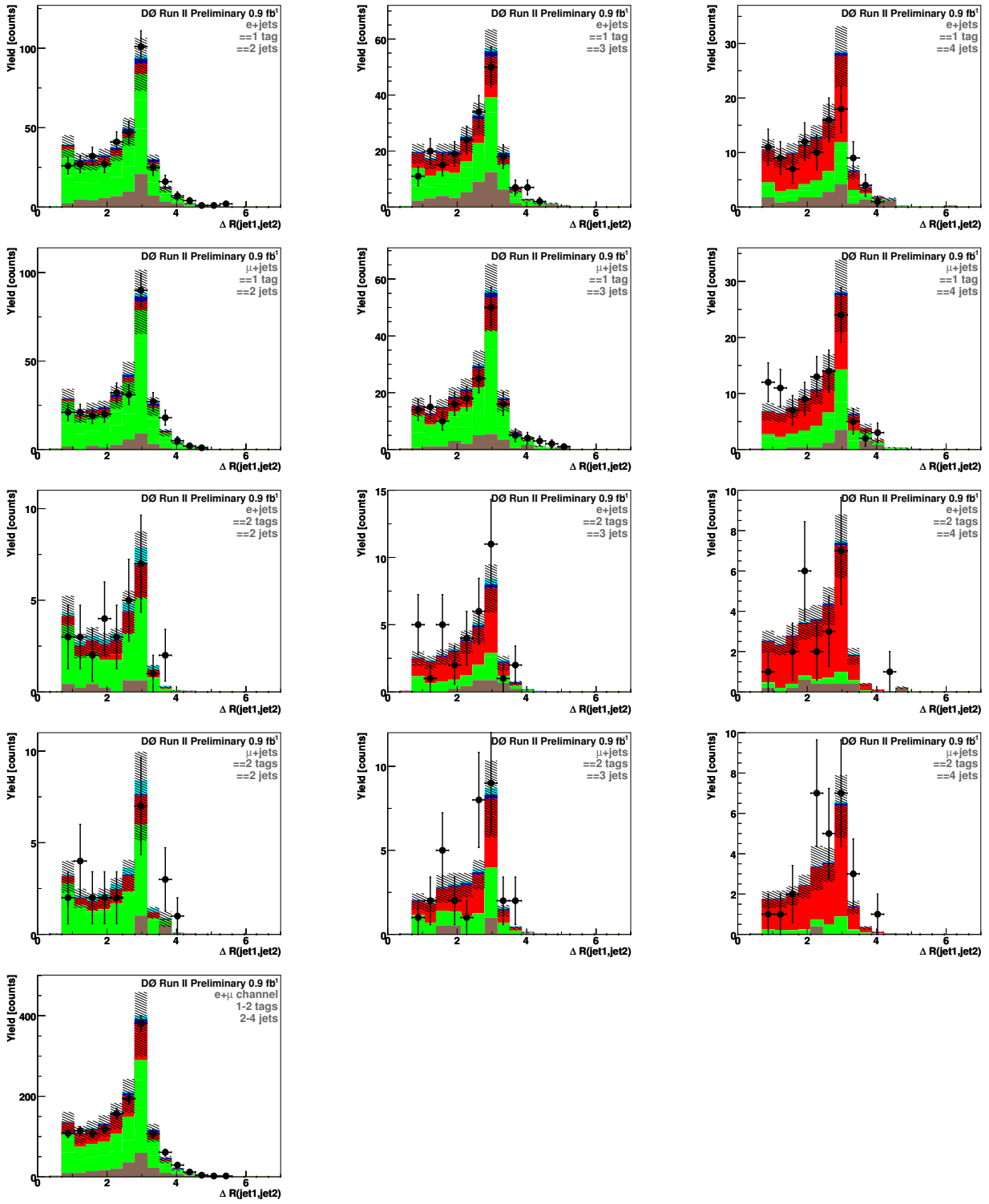


Figure C.5: The opening angle $\Delta R(\text{jet1}, \text{jet2})$ for events with two jets (left column), three jets (center column), and three jets (right column), for electron/1tag (first row), muon/1tag (second row), electron/2tags (third row), and muon/2tags (fourth row). The plot at the bottom of the figure shows the distribution for the electron and muon channels combined, for 2, 3, and 4 jets combined, and for 1 and 2 tags combined, i.e., all the channels we have used for this analysis.

REFERENCES

[1] M. Veltman, *Facts and Mysteries in Elementary Particle Physics,* (World Scientific, 2003). 1.1

[2] D. Griffiths, *Introduction to Elementary Particles,* (John Wiley Sons, 1987). 1.1

[3] F. Halzen and A. Martin, *Quarks and Leptons,* (John Wiley Sons, 1984). 1.1, 1.1.2

[4] C. Quigg, *Gauge Theories of the Strong, Weak, and Electromagnetic Interactions,* (Westview Press, 1983). 1.1

[5] R. Ellis *et al.*, *QCD and Collider Physics,* (Cambridge University Press, 1996). 1.1

[6] H. B. Prosper and M. Danilov, *Techniques and Concepts of High-Energy Physics XII,* (NATO Science Series, 2003). 1.1

[7] S. Eidelman *et al.*, "Review of Particle Physics," Phys. Lett. B, 592:1, (2004). 1.1.2

[8] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory,* (Harper-Collins Publishers, 1995). 1.1.2

[9] E. Gross *et al.*, Z. Phys. **C63**, 417 (1994); Erratum: *ibid.*, **C66**, 32 (1995). 1.1.4

[10] A. Djouadi, M. Spira, and P.M. Zerwas, Z. Phys. **C70**, 675 (1996). 1.1.4

[11] ALEPH, DELPHI, L3 and OPAL Collaborations, Phys. Lett. B **565**, 61(2003) 1.1.4

[12] N. Cabibbo *et al.*, Nucl. Phys. **B158**, 295 (1979); T. Hambye and K. Riesselmann, Phys. Rev. **D55**, 7255 (1997); G. Isidori *et al.*, Nucl. Phys. **B609**, 387 (2001). 1.1.4

[13] B.W. Lee *et al.*, Phys. Rev. Lett. **38** (1977) 883; M. Quiros, *Constraints on the Higgs boson properties from the effective potential*, hep- ph/9703412; A. Ghinculov and T. Binoth, Acta Phys. Polon. **B30** (1999) 99. 1.1.4

[14] L. Maiani, G. Parisi and R. Petronzio, Nucl. Phys. **B136** (1979) 115; N. Cabibbo *et al.*, Nucl. Phys. **B158** (1979) 295; R. Dashen and H. Neunberger, Phys. Rev. Lett. **50** (1983) 1897; D.J.E. Callaway, Nucl. Phys. **B233** (1984) 189; M.A. Beg *et al.*, Phys. Rev. Lett. **52** (1984) 883; M. Lindner, Z. Phys. **C31** (1986) 295. 1.1.4

[15] G. Altarelli and G. Isidori, Phys. Lett. **B337** (1994) 141; J.A. Casas, J.R. Espinosa and M. Quiros, Phys. Lett. **B342** (1995) 171, Phys. Lett. **B383** (1996) 374; B. Grzadkowski and M. Lindner, Phys. Lett. **B178** (1986) 81; T. Hambye and K. Riesselmann, Phys. Rev. **D55** (1997) 7255. 1.1.4

[16] N. Varelas, SM Higgs Searches at the Tevatron, HEP- EPS Conference, Lisbon, July 21-27, 2005. 1.1.4

[17] F. Abe *et al.*, "Observation of top quark production in $p\overline{p}$ collisions," Phys. Rev. Lett. 74:2626-2631 (1995) 1.2

[18] S. Abachi *et al.*, "Observation of the top quark," Phys. Rev. Lett. 74:2632-2637 (1995) 1.2

[19] I. I. Y. Bigi *et al.*, "Production and decay properties of ultraheavy quarks," Phys. Lett. B181:157 (1986) 1.2

[20] E. Brubaker *et al.*, "Combination of cdf and dØ results on the mass of the top quark," FERMILAB-TM-2355-E (2006). 1.2, 1.2

[21] S. Abachi *et al.*, DØCollaboration, Phys. Rev. Lett. **79**, 1203 (1997) 1.2

[22] F. Abe *et al.*, CDF Collaboration, Phys. Rev. Lett. **80**, 273 (1998) 1.2

[23] B. Abbott *et al.*, DØCollaboration, Phys. Rev. **D60**, 052003 (1999) 1.2

[24] F. Abe *et al.*, CDF Collaboration, Phys. Rev. Lett. **82**, 271 (1999) 1.2

[25] V. M. Abazov *et al.*, "Measurement of the ttbar production cross section in ppbar collisions at $\sqrt{s} = 1.96$ TeV using kinematic characteristics of lepton + jets events," Phys. Rev. D, (Submitted on 19 May 2007). 1.2

[26] CDF Collaboration, "CDF II Top Quark Group Physics Results," 1.2
http://www-cdf.fnal.gov/physics/new/top/top.html.

[27] T. M. P. Tait and C. P. Yuan, "Single top quark production as a window to physics beyond the standard model," Phys. Rev. D 63(1)014018 (2000). 1.3.1

[28] T. Affolder *et al.*, "First Measurement of the Ratio $B(t \to Wb)/B(t \to Wq)$ and Associated Limit on the CKM Element V_{tb}," Phys. Rev. Lett. 86, 3233 (2001) 1.3.1

[29] G. Mahlon and S.J. Parke, "Improved Spin Basis for Angular Correlation Studies in Single Top Quark Production at the Tevatron," Phys. Rev. D 55, 7249 (1997); G. Mahlon and S.J. Parke, "Single Top Quark Production at the LHC: Understanding Spin," Phys. Lett. B 476, 323 (2000). 1.3.1

[30] E.E. Boos and A.V. Sherstnev, "Spin Effects in Processes of Single Top Quark Production at Hadron Colliders," Phys. Lett. **B534**, 97 (2002). 1.3.1

[31] A. Czarnecki and M. Jezabek, "Distributions of Leptons in Decays of Polarized Heavy Quarks," Nucl. Phys. **B427**, 3 (1994). 1.3.1

[32] A. Czarnecki, M. Jezabek and J.H. Kuhn, "Lepton Spectra From Decays Of Polarized Top Quarks," Nucl. Phys. **B351**, 70 (1991). 1.3.1

[33] A. Brandenburg, Z.G. Si and P. Uwer, "QCD-Corrected Spin Analyzing Power of Jets in Decays of Polarized Top Quarks," Phys. Lett. **B539**, 235 (2002). 1.3.1

[34] A.P. Heinson, A.S. Belyaev and E.E. Boos, "Single Top Quarks at the Fermilab Tevatron," Phys. Rev. D **56**, 3114 (1997). 1.3.1

[35] M. Agelou *et al.*, "Search for Single Top Quark Production at DØ in Run II," DØNote 4398 (2004). 1.3.1

[36] J. Thompson, "Introduction to Colliding Beams at Fermilab," FERMILAB-TM-1909 (1994). 2.1

[37] General Reference Documents for Accelerators. 2002. 2.1

http://www-numi.fnal.gov/workgrps/protonwg/accel_reference.html.

[38] Accelerator Concepts Rookie Books. 2003. 2.1

http://www-bdnew.fnal.gov/operations/rookie_books/rbooks.html.

[39] D. Edmunds, "Run IIa Tevatron Beam Structure," 2.1

http://www.pa.msu.edu/hep/d0/ftp/l1/framework/drawings/run_iia_beam_structure.ps.

[40] T. Ferbel, "A Brief Description of the DØDetector in Run II," 2.2

http://www-d0.fnal.gov/Run2Physics/WWW/templates/detector.tex.

[41] S. Abachi *et el.*, The DØDetector. Nucl. Instrum. Methods Phys. Res. A , 338:185, 1994. 2.2

[42] DØCollaboration, "DØSilicon Tracker Technical Design Report," 2.2.2

http://d0server1.fnal.gov/projects/silicon/www/tdr_final.ps.

[43] DØCollaboration, "The DØUpgrade: Central Tracker Technical Design Report," 2.2.2

http://d0server1.fnal.gov/users/stefan/www/CFT_TDR/CFT_TDR.ps.

[44] M. Adams *et al.*, "Design Report of the Central Preshower Detector for the DØUpgrade," 2.2.3

http://d0server1.fnal.gov/users/qianj/CPS/doc/dn3104.pdf.

[45] A. Gordeev *et al.*, "Technical Design Report of the Forward Preshower Detector for the DØUpgrade," DØNote 3445 (1998). 2.2.3

[46] B. Baldin *et al.*, "Technical Design Report of the Central Muon System," DØNote 3365 (1997). 2.2.5

[47] G. Alexeev *et al.*, "Technical Design Report of the DØForward Muon Tracking Detector Based on Mini-Drift Tubes," DØNote 3366 (1997). 2.2.5

[48] A. Khanov, "HTF: histogramming method for finding tracks. The algorithm description," DØNote 3778 (2000). 3.1

[49] C. Tully and A. Schwartzman, "Primary vertex reconstruction by means of adaptive vertex fitting," DØNote 4918 (2005). 3.2

[50] J. Hays *et al.*, "Single Electron Efficiencies in p17 Data and Monte-Carlo," DØNote 5025 (2006). 3.3

[51] A. Kumar *et al.*, "Electron Likelihood Study," DØNote 4769 (2005). 3.3

[52] P. Calfayan *et al.*, "Muon Identification Certification for p17 data," DØNote 5157 (2006). 3.4

[53] V. M. Abazov *et al.*, "The Upgraded DØ Detector," Nucl. Instrum. Meth. **A**, (2006). 3.4, 3.5, 5.4

[54] G. C. Blazey *et al.*, "Run II Jet Physics," DØNote 3750 (2000). 3.5

[55] DZero Preliminary Jet Energy Scale: 3.5
http://www-d0.fnal.gov/phys_id/jes/public/plots_v7.1.

[56] T. Scanlon *et al.*, "Performance of the NN b-tagging Tool on p17 Data," DØNote 5213 (2006). 3.6

[57] E.E. Boos *et al.*, "Method for Simulating Electroweak Top-Quark Production Events in the NLO Approximation: SingleTop Generator," Phys. Atom. Nucl. **69**, 1317 (2006). 5.1

[58] Sjoestrand, T. et al.: PYTHIA 6.3 Physics and Manual. hep-ph/0308153 , 2003. 5.1

[59] S. Jadach *et al.*, "The Tau Decay Library TAUOLA: version 2.4," Comput. Phys. Commun. **76**, 361 (1993). We used version 2.5. 5.1

[60] D.J. Lange, "The EvtGen Particle Decay Simulation Package," Nucl. Instrum. Meth. **A 462**, 152 (2001). We used version 00.00.17. 5.1

[61] M.L. Mangano *et al.*, "ALPGEN, a Generator for Hard Multiparton Processes in Hadronic Collisions," J. High Energy Phys. **0307**, 001 (2003). We used ALPGEN version 2.05. 5.3.1

[62] S. Höche *et al.*, "Matching Parton Showers and Matrix Elements," hep-ph/0602031. 5.3.2

[63] M. Agelou *et al.*, "Top Trigger Efficiency Measurements and the top_trigger Package," DØNote 4512 (2004). 5.4

[64] Y. Fisyak and J. Womersley, "DØgstar DØGEANT Simulation o fthe Total Apparatus Response," DØNote 3191 (1997). 5.5

[65] Data Quality Group web page: http://www-d0.fnal.gov/computing/data_quality/d0_private/forusers.html. 6

[66] C. M. Bishop, *Neural Networks for Pattern Recognition,* (Clarendon Press, Oxford, 1998). 7, 7.1, 8.3.1

[67] D. S. Sivia, *Data Analysis: A Bayesian Tutorial,* (Oxford Science Publications, 1996) 7

[68] R. M. Neal, *Bayesian Learning of Neural Networks,* (Springer-Verlag, New York, 1996). 7, 7.4.4

[69] R. T. Cox, "Probability, frequency and reasonable expectation," American Jornal of Physics 14(1):1-13, (1946). 7.2

[70] L. Paninski, "Markov Chain Monte Carlo and Gibbs Sampling," 7.4.1

(http://www.stat.columbia.edu/l̃iam/teaching/neurostat-spr07/papers/mcmc/mcmc-gibbs-intro.pdf).

[71] R. M. Neal, "Probabilistic Inference Using Markov Chain Monte Carlo Methods," 7.4.1

http://omega.albany.edu:8008/neal.pdf.

[72] W. K. Hastings, *Monte Carlo sampling methods using Markov chains and their applications,* (Biometrika, 1970). 7.4.2, 7.4.2

[73] S. Geman and D. Geman, *Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images,* (IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 6, 1984). 7.4.3

[74] A. E. Gelfand and A. F. M. Smith, *Sampling-based approaches to calculating marginal densities,* (Journal of the American Statistical Association, vol. 85, 1990). 7.4.3

[75] S. Duane, A. D. Kennedy, B. J. Pendleton and D. Roweth, "Hybrid Markov Chain Monte Carlo," Phys. Lett. **B** 195, 216-222 (1987). 7.4.4

[76] R. Barlow, "Event Classification Using Weighting Methods," J. Comp. Phys. **72**, 202 (1987). 8.1

[77] The Single Top Working Group, "Using Bayesian Neural Networks to Search for Single Top Quarks in 1 fb^{-1} of Data", DØNote 5288 (2006). 8.1.1

[78] J. H. Friedman, "Separating Signal From Background Using Ensembles Of Rules," in *Statistical Problems in Particle Physics, Astrophysics and Cosmology,* ed. L. Lyons and M. K. Ünel, (Imperial College Press, London, 2006); 2, 8.2.2

http://www-stat.stanford.edu/∼jhf/R-RuleFit.html.

[79] L. Dudko, "Use of Neural Networks in a Search for Single Top Quark Production at DØ," AIP Conf. Proc. **583**, 83 (2001). 8.2

[80] E. Boos and L. Dudko, "Optimized Neural Networks to Search for Higgs Boson Production at the Tevatron," Nucl. Instrum. Meth. **A 502**, 486 (2003). 8.2

[81] Q.H. Cao, R. Schwienhorst, and C.P. Yuan, "Next-to-Leading Order Corrections to Single Top Quark Production and Decay at the Tevatron: 1: s-Channel Process," Phys. Rev. D **71**, 054023 (2005). 8.2

[82] C. Peterson, T. Rögnvaldsson, and L. Lönnblad, *JETNET 3.0-A versatile artificial neural network package,* Computer Physics Communications **81**, 185-220 (1994). 8.3.1

[83] E. Aguilo *et al.*, "A Second Look at Bayesian Neural Networks in the Search for Single Top Quarks in in 1 fb^{-1} of Data," DØNote 5361 (2007). 8.3.2

[84] S. Jain *et al.*, "Computing Limits Using a Bayesian Approach in the Package `top_statistics`," DØ Note 5123 (2006). 8.6

[85] E. Aguilo *et al.*, "Search for Single Top Quark Production in 1 fb^{-1} of Data," DØNote 5285 (2006). 8.6.1

BIOGRAPHICAL SKETCH

Fermi National Accelerator Laboratory
MS 352 Batavia, IL 60510
Tel: (630) 840-8384 Fax: (630) 840-8886
Email: dkau@fnal.gov

EDUCATION

August 1999 -- Present ***Ph.D.*** **Physics** ***Florida State University, Tallahassee, FL, USA***
- *Dissertation: Evidence for the single top quark production using Bayesian neural networks*
- *Advisor: Dr. Harrison B. Prosper*

March 1992 – February 1998 ***B.S.*** **Physics** ***Kyung Won University, Seongnam, Gyeonggi-do, Republic of Korea***

RESEARCH EXPERIENCE

January 2002 - Present **Research Assistant**

Fermi National Accelerator Laboratory (FNAL), Batavia, IL
Department of Physics, Florida State University, Tallahassee, FL

High Energy Physics Data Analysis

- Developed analysis frameworks using C++, Python and shell scripts
- Generated and analyzed Monte Carlo simulation (MC) data to estimate signal and background components
- Employed and studied various statistical approaches to test the agreement between MC simulated data and the observed data
- Bayesian neural network analysis for my Ph.D. dissertation
  - Investigated neural network techniques and underlying theories in detail
  - Developed codes to support the operation of the neural network software
- Bayesian statistics to calculate the cross section of the single top quark production with systematic uncertainties
- Performed pseudo experiments or ensemble tests to validate the results

Collider Detector R&D

- Performance tests of SVX4 chips which are analog to digital converter chips used by DØ and CDF (collider detector experiments at the FNAL)
  - Documented in Fermilab technical memo **FERMILAB-TM-2317-E, Jun 2005. 31pp.**
- Investigation of electronic properties and noise features of layer0 a silicon detector installed in DØ spring 2006

May 2000 – December 2001 **Research Assistant**

National High Magnetic Field Laboratory (NHMFL), Tallahassee, FL
Department of Physics, Florida State University, Tallahassee, FL

Condensed Matter Physics

- Development of the femtosecond electron diffraction system
  - Assembled a vacuum chamber capable of reaching ~ 10^{-11} torr
  - Developed an imaging system using a CCD camera, a P11 phosphor screen and a Chevron-type microchannel plate
  - Operated a Ti:sapphire laser system which provides sub-50 fs pulses and wavelength of ~800 nm
  - Built an electron gun, a magnetic lens and electronic devices to synchronize the laser and the CCD camera
  - Designed and set up configurations of optical components (lenses, filters, beamsplitters etc) for laser beam control
  - Worked on growth of thin metal films using an evaporator
  - Performed a streak camera experiment to measure the pulse duration of electron beams
  - Analyzed electron diffraction images of metal films (Au, Ag, Al etc)

TEACHING EXPERIENCE

August 1999 – April 2000 **Teaching Assistant**

Department of Physics, Florida State University, Tallahassee

SKILLS

- Programming Language: C, C++, Python, Visual Basic
- Operating System: Unix, Linux, Windows
- Software: Latex, ROOT, MS Office, Maple, R (a language and environment for statistical computing and graphics), neural network programs (Software for Flexible Bayesian Modeling, JETNET, MLPfit etc)
- Hardware: Evaporator, Lathe, Ultra-High Vacuum Chamber, Optical Instruments, CCD Imaging System, Femtosecond Electron Diffraction System
- Spoken Languages: Korean, English

PROFESSIONAL ORGANIZATIONS

American Physical Society

PRESENTATIONS

[1] Daekwang Kau, *Improved Search for Single Top Quark Production Using Bayesian Neural Networks* (Oral Presentation), ***PHENO Symposium, Madison***, WI, 2007

[2] Daekwang Kau, *Measurement of B(t->Wb)/B(t->Wq) at DØ* (Oral Presentation), ***DPF Meeting, Honolulu***, HI, 2006

[3] Daekwang Kau, *Search for Electroweak Production of the Top Quark in the Muon+Jets Channel at DØ* (Oral Presentation), ***APS April Meeting, Tampa***, FL, 2005

[4] Daekwang Kau, *Design and Performance Studies of a Layer0 Detector at DØ* (Oral Presentation), ***APS April Meeting, Denver***, CO, 2004

[5] Daekwang Kau, *The Layer0 Silicon Detector*, ***DOE 2004 Annual Fermilab Program Review Poster Session***, IL, 2004

[6] Hyuk Park, Zhao Hao, Daekwang Kau, Chenggang Tao, Lukasz Blaszczyk, Jianming Cao, *A Femtosecond Electron Diffractometer for the Study of Ultrafast Structural Dynamics* (Oral Presentation), ***APS March Meeting, Austin***, TX, 2003

PRIMARY AUTHORSHIP

[1] The DØ Collaboration, *Evidence for Production of Single Top Quarks and first direct measurement of* $|V_{tb}|$, published in ***Phys. Rev. Lett. (2006).***

[2] The DØ Collaboration, *Multivariate Searches for Single Top Quark Production with the DØ Detector*, published in ***Phys. Rev. D (2006).***

[3] The DØ Collaboration, *Search for Single Top Quark Production in* $p\bar{p}$ *Collisions at* $\sqrt{s}$ *=1.96 TeV*, published in ***Phys. Lett. B (2005).***

[4] Jianming Cao, Zhao Hao, Hyuk Park, Chenggang Tao, Daekwang Kau, Lukasz Blaszczyk, *Femtosecond Electron Diffraction for Direct Measurement of Ultrafast Atomic Motions*, published in ***Appl. Phys. Lett. (2003).***

REFERENCES

Professor Harrison B. Prosper
Department of Physics, Florida State University
514 Keen Building
Tallahassee, FL 32306-3016
Phone: (850) 644-6760
Fax: (850) 644-6735
E-mail: harry@hep.fsu.edu

Professor Reinhard Schwienhorst
Physics and Astronomy Department, Michigan State University
3234 Biomedical and Physical Sciences Building
East Lansing, MI 48824-2320
Phone: (517) 355-9200, ext. 2136
Fax: (517) 353-4500
E-mail: schwier@pa.msu.edu

Professor Kazunori Hanagaki
Department of Physics, Osaka University
1.1 Machikaneyama, Toyonaka
Osaka 560-0043, Japan
Phone: (81-6) 6842-2560
Fax: (81-6) 6850-5532
E-mail: kazu@fnal.gov